Institut für Parasitologie
der Vetsuisse-Fakultät Universität Zürich

Direktor: Prof. Dr. med. vet. Peter Deplazes
Leiter Abteilung Vektor Entomologie: Prof. Dr. sc. nat. ETH Alexander Mathis

Arbeit unter wissenschaftlicher Betreuung von
Prof. Dr. med. vet. Cornelia Silaghi

De novo genome sequencing and comparative stage-specific transcriptomic analysis of *Dirofilaria repens*

Inaugural-Dissertation

zur Erlangung der Doktorwürde der
Vetsuisse-Fakultät Universität Zürich

vorgelegt von

Chiara Valentina Cafarelli

Tierärztin
von Zweisimmen, Bern

genehmigt auf Antrag von

Prof. Dr. sc. nat. ETH Alexander Mathis, Referent

2019

Inhaltsverzeichnis

Vetsuisse-Faculty University of Zurich (2019)

Chiara Valentina Cafarelli

Institute of Parasitology, parasito@vetparas.uzh.ch

De novo genome sequencing and comparative stage-specific transcriptomic analysis of *Dirofilaria repens*

The zoonotic mosquito-borne filarial nematode *Dirofilaria repens* causes subcutaneous and ocular infections in dogs, cats and humans. *D. repens* is transmitted by mosquitoes by ingesting microfilariae (mf) from an infected host, which develop in the mosquito to the infectious third stage larvae (L3). The aims of the project were the *de novo* sequencing and annotation of the *D. repens* genome and comparative transcriptomic analyses of the developmental stages mf and L3. The 99.59 MB genome was around 17% larger than that of the related species *D. immitis* and contained 8.9% fewer predicted genes (10,357). A significantly higher number of *D. repens* proteins as compared to *D. immitis* mapped to the filarial nematode *L. loa*, reflecting the similarity in biology of *D. repens* and *L. loa*. A total of 876 genes were differentially expressed, of which 591 could be annotated in UniProtKB/Swiss-Prot. In particular, 155 genes with a UniProtKB/Swiss-Prot annotation to *C. elegans* and filarial nematodes were upregulated in the L3 and 57 in the mf stage, respectively. Fifteen Gene Ontology Biological Processes were significantly enriched for the L3 group and 12 for the mf. These data provide first insight into the differential gene expression profiles of this filarial nematode and can serve for future investigations of metabolic processes and stage-specific diagnostics.

Keywords: *Dirofilaria repens*, genome, transcriptome, microfilariae, third stage larvae.

De-novo Genomsequenzierung und vergleichende stadienspezifische Transkriptomanalyse von *Dirofilaria repens*

Der zoonotische Nematode *Dirofilaria repens* verursacht subkutane und okulare Infektionen bei Hunden, Katzen und Menschen. *D. repens* wird durch Stechmücken übertragen, welche Mikrofilarien (Mf) aus dem Blutstrom eines infizierten Wirtes aufnehmen. Mf entwickeln sich in der Mücke zu infektiösen Drittlarven (L3). Die Ziele des Projekts waren die *De-novo* Sequenzierung und Annotation des *D. repens*-Genoms, sowie vergleichende Transkriptomanalysen der Entwicklungsstadien Mf und L3. Das 99,59 MB große Genom war rund 17% größer als das der verwandten Spezies *D. immitis* und enthielt 8,9% weniger vorhergesagte Gene (10.357). Eine wesentlich höhere Anzahl von *D. repens* Proteinen zeigte eine Ähnlichkeit zum Nematoden *L. loa* verglichen mit *D. immitis*, was die Ähnlichkeit der Biologie von *D. repens* und *L. loa* widerspiegelt. Insgesamt wurden 876 Gene differentiell exprimiert. Davon konnten 591 in UniProtKB/Swiss-Prot annotiert werden. Insbesondere wurden 155 Gene mit einer UniProtKB/Swiss-Prot-Annotation für *C. elegans* und filariale Nematoden im Stadium L3 bzw. 57 im Stadium mf hochreguliert. Fünfzehn „Gene Ontology Biological Processes“ wurden für die L3 und 12 für die Mf signifikant angereichert. Diese Daten geben einen ersten Einblick in die differenziellen Genexpressionsprofile von *D. repens* und können für zukünftige Untersuchungen von Stoffwechselprozessen und stadienspezifischen Diagnostika verwendet werden.

Stichworte: *Dirofilaria repens*, Genom, Transkriptom, Mikrofilarien, Drittlarven

2. Manuskript

De novo genome sequencing and comparative stage-specific transcriptomic analysis of *Dirofilaria repens*

Chiara Cafarelli[a,*], Giancarlo Russo[b,*], Alexander Mathis[a], Cornelia Silaghi[a,c**]

[a]National Centre for Vector Entomology, Institute of Parasitology, Vetsuisse Faculty, University of Zürich, Winterthurerstr 266a, 8057 Zürich, Switzerland; [b]Functional Genomics Center Zürich, University of Zürich and ETH Zürich, Winterthurerstr. 190, 8057 Zürich, Switzerland; [c]Friedrich-Loeffler-Institute, Federal Research Institute for Animal Health, Südufer 10, 17493 Greifswald - Insel Riems, Germany.

*both authors contributed equally to this manuscript and share the first authorship

**corresponding author

E-mail address: Cornelia.Silaghi@fli.de

Friedrich-Loeffler-Institute, Federal Research Institute for Animal Health, Südufer 10, 17493 Greifswald - Insel Riems, Germany

2.1 Abstract

The zoonotic mosquito-borne filarial nematode *Dirofilaria repens* causes subcutaneous and ocular infections in dogs, cats and humans. Microfilariae (mf) are taken up by mosquitoes from infected vertebrate hosts which develop in the mosquito to the infectious third stage larvae (L3). These are transmitted to new vertebrate hosts and develop over two further moults to adult worms. The aims of the project were 1) the *de novo* sequencing and annotation of the *D. repens* genome and 2) comparative transcriptomic analyses of the developmental stages mf and L3. Genomic DNA was obtained from adult male *D. repens*. RNA was extracted from microfilariae from naturally infected dogs and from L3 produced in *Aedes aegypti* mosquitoes fed on blood spiked with microfilariae. The 99.59 MB genome was around 17% larger than that of the related species *D. immitis* (dog heartworm) and contained 8.9% fewer predicted genes (10,357). Around 1.8% of identified proteins (206/11,262) could not be mapped to *D. immitis*. Out of these, 6 (2.9%) presented an ortholog in all other considered filarial nematodes (e.g. *Loa loa*) and *Caenorhabditis elegans*. A significantly higher number of *D. repens* proteins as compared to *D. immitis* mapped to the filarial nematode *L. loa*, reflecting the similarity in biology of *D. repens* and *L. loa*. A total of 876 genes were differentially expressed, of which 591 could be annotated in UniProtKB/Swiss-Prot. In particular, 155 genes with a UniProtKB/Swiss-Prot annotation to *C. elegans* and filarial nematodes were upregulated in the L3 and 57 in the mf stage, respectively. Fifteen GO BP (Gene Ontology Biological Processes) were significantly enriched for the L3 group and 12 for the mf. These data provide first insight into the differential gene expression profiles of this filarial nematode and can serve for future investigations of metabolic processes and stage-specific diagnostics.

Keywords: *Dirofilaria repens*, genome, transcriptome, microfilariae, third stage larvae.

2.2 Introduction

Dirofilaria repens is a mosquito-borne filarial nematode of the Onchocercidae family causing subcutaneous and ocular infections in dogs and cats. The life cycle of *D. repens* contains five developmental stages and requires both an arthropod vector (mosquito species mostly of the genera *Culex, Aedes and Anopheles*) which are also the intermediate host (Simon et al., 2012; Silaghi et al., 2017) and a mammalian definitive host. After mating, adult female *D. repens* worms release microfilariae (mf) into the blood stream of the mammalian host. Circulating mf (L1) are taken up by a bloodsucking mosquito in which the mf invade after approximately 24h the cells of the Malpighian tubules, develop into the so-called sausage form, then undergo two moults to second stage (L2) and finally to the infective third stage larvae (L3). For a successful transmission, the L3 have to migrate from the Malpighian tubules to the proboscis of the mosquito. During a subsequent blood meal, infectious L3 actively leave the labiae and penetrate the host's skin (Simon et al., 2012) where they moult to the fourth larval stage (L4). These stages move to the subcutaneous tissue and muscular connective fasciae where they develop after a last moult to the adult stage and reside permanently. Adult worms live up to 4 years in their natural hosts (Genchi and Kramer, 2017). Dogs are the main reservoir, while wild carnivores or cats are rarely microfilaraemic (Magi et al., 2008). Furthermore, humans can be accidental hosts for *D. repens*. Subcutaneous or the often clinically more severe ocular dirofilariosis is considered an emerging zoonosis in Europe and is already well-known in Eastern countries such as southern Russia and Ukraine (Capelli et al., 2018). *Dirofilaria repens* occurs in various regions of Europe, Africa, and Asia. In recent years, *D. repens* has been expanding northwards, assumedly partly due to an increasing number of dogs travelling within Europe, and climate change (Genchi et al., 2011). As is the case in many filarial nematodes, *D. repens* harbors intracellular symbiotic bacteria of the genus *Wolbachia* which are thought to be beneficial to the nematode host (Taylor et al., 2005; Slatko et al., 2010; Godel et al., 2012).

In contrast to the related filarial nematode *D. immitis* (heartworm), for which the genome and few transcriptome studies have recently been published (Chaisson and Tesler, 2012; Godel et al., 2012; Luck et al., 2014), neither data on the nuclear genome nor on the transcriptome of *D. repens* is yet available.

Therefore, the aim of the present study was threefold: first, we aimed to fill the gap in the genomics landscape of filarial nematodes by generating the first comprehensive resource for the organism *D. repens;* second, we tried to investigate whether differences in the genomes and the transcriptomes of the filarial nematodes, and in the particular between the two *Dirofilariae*, could point to specific traits of these organisms; finally, we wanted to identify which genes contribute the most to the transcriptional activities in the different stages analysed, i.e., mf in the mammalian host and L3 in the mosquitoes. Differences in the overall transcriptional profiles between these groups are to be expected, however, the main focus was to identify stage-specific genes which would help to gain first insights into metabolic pathways of this filarial nematode.

2.3 Material and Methods

2.3.1 Parasite material

The adult male *D. repens* worm used in this study for DNA sequencing originated from an experimental study in Italy (Genchi et al., 2013). It had been collected from a beagle dog (ID 9478-9800) during necropsy and was continuously stored at -20 °C.

EDTA-anticoagulated blood samples were obtained from dogs naturally infected with *D. repens*. During routine diagnosis, three dogs from a shelter in Lithuania were found to have microfilariaemic blood which was used for RNA sequencing from microfilariae, whereas microfilariaemic blood of three privately-owned dogs diagnosed in Croatia was used to produce L3

of *D. repens* in mosquitoes. The fresh blood samples of Lithuania were processed on site up to the worm homogenization step (see below). The samples were stored at -40 °C overnight and then shipped with a courier to the Institute in Zürich. Immediately after arrival, they were put at -80 °C and further processed the following day. EDTA-anticoagulated blood samples from Croatia were placed in a polystyrene box containing cooling elements to keep microfilariae alive and sent the next day with a courier to Zürich.

Vitality and number of microfilariae were confirmed by microscopy as described in a previous study with the modification that microfilariaemiae was calculated as average from four counts (Silaghi et al., 2017).

The infections with *D. repens* and the absence of *D. immitis* were confirmed by traditional PCRs as described (Rishniw et al., 2006). PCRs were run on the PTC-200 Peltier thermal cycler (Bio-Rad, Reinach, Switzerland), and the amplicons examined on a 1.5% agarose gel under UV light.

2.3.2 Mosquitoes and production of L3

A laboratory colony of *Aedes aegypti* was maintained in a climate chamber in an insectarium under standard laboratory conditions at a temperature of 27 °C, a relative humidity (rh) of 85% and a light-dark cycle of 16:8 h including dusk/dawn phases of 1 h. Mosquitoes were kept in purpose-built plastic boxes (dimensions: 32.7 cm x 16.3 cm x 22.7 cm, volume 8.2 liters) containing racks of plates to increase the surface. In average, each box contained 4000 mosquitoes. Sugar cubes and water ad libitum were provided. For reproduction, the mosquitoes were fed with EDTA-anticoagulated bovine blood from the local slaughterhouse using a standard artificial feeding system (Hemotek, Hemotek Ltd, Lancashire, UK).

Mosquitoes were fed at room temperature for 2 hours through stretched Parafilm membranes in a Hemotek system on 3 ml spiked blood at 37 °C (Silaghi et al., 2017). Three biological

replicates were done with blood samples containing 4675, 4000 or 150 mf/ml, respectively. After the exposure to the blood meal, all mosquitoes (fed and unfed) were kept at the standard conditions mentioned above. After 14 days, the mosquitoes were anaesthetized with diethylether (≥ 95.5%) and gently ground with a mortar and a pestle in order to enable the L3 to escape the mosquitoes' bodies without damaging them. The material was then rinsed onto a 100 µm mesh sieve placed in a glass bowl (22 cm diameter) containing Hank's balanced salt solution (HBBS) (Thermo Fisher, Waltham, USA) for 2.5 h at 27 °C and 85% rh. The L3 migrated through the sieve into the warm HBSS to the bottom of the glass bowl. The L3 were then filtered through two sieves (mesh sizes 32 µm, 22 µm) and single L3 picked using a binocular and put into RNAlater (Thermo Fisher) and stored at -20 ºC overnight.

2.3.3 Nucleic acid extraction

The *D. repens* worm was placed in a 2 ml Eppendorf tube with 500 µl Tris-EDTA buffer and disrupted with a Tissue Lyser II (Qiagen, Hilden, Germany) with one 5 mm stainless steel bead at 30 beats per second for 1 min. This step was repeated twice. In between the disruption processes, the sample was centrifuged at room temperature for 30 s (10'000 x g). Genomic DNA was extracted using the Gentra Purogene Tissue Kit (Qiagen) according to the manufacturer's instructions using the tissue protocol for processing 50-100 mg tissue. The sample was stored at -20 °C until further processing.

A total of 4 - 6 ml of microfilariaemic blood was drawn up into a 10 ml syringe and immediately diluted and lysed in the syringe with diethyl pyrocarbonate (DEPC)-treated distilled water (prewarmed to 37 °C). The solution was injected through the Difil-filter system (mesh size 5 µm) (Evsco, Buena, USA) and thoroughly washed with DEPC-treated water (37 °C). The filter was transferred to a 2 ml Eppendorf tube filled with 1 ml of Trizol reagent (Invitrogen, Carlsbad, USA) and cut into small pieces with sterile scissors. Five cycles of flash-freezing in liquid N_2

and crushing with plastic pestles were performed to obtain homogeneous worm extracts according to a method described previously (Ballesteros et al., 2016). RNA was isolated by extraction with the Trizol LS kit (Invitrogen), followed by column purification using the RNeasy mini kit (Qiagen) according to the manufacturer's instructions. Briefly, the homogenized samples were incubated for 5 min at room temperature to permit the complete dissociation of nucleoprotein complexes. Two hundred µl of chloroform was added to each tube and the samples were vortexed and incubated for 3 min at room temperature. The samples were then centrifuged at 12'000 x g at 4 °C for 15 min. The aqueous phase was transferred to fresh tubes and mixed with an equal volume of cold ethanol (70%). The tubes were vortexed for 5 sec and the mixture transferred to an RNA binding spin column from the RNeasy kit (Qiagen). In order to eliminate genomic DNA contamination, the samples were treated with DNase using an RNase-Free DNase Set DNA-free Kit (Qiagen).

L3 were concentrated by filtration (Difil test) and washed with DEPC-treated distilled water. The filters were transferred to 2 ml Eppendorf tubes containing 1 ml TRIZOL and cut into small pieces with sterile scissors. The samples were disrupted in a Tissue Lyser II (Qiagen) with 15-20 glass beads (ø 2.85 – 3.45 mm) at 30 beats per second for 1 min. RNA was isolated as described for microfilariae.

2.3.4 DNA sequencing

SMRT bell templates were produced using the DNA Template Prep Kit 1.0 (Pacific Biosciences, Menlo Park, USA) (p/n 100-259-100). The input genomic DNA concentration was measured using a Qubit Fluorometer dsDNA Broad Range assay (Thermo Fischer) (p/n 32850), revealing a concentration of 88 ng/µl (total amount 25.2 µg). A total of 10 µg of gDNA was mechanically sheared to an average size distribution of 15 kbp, using a Megaruptor Device (Diagenod, Seraing, Belgium). A Bioanalyzer 2100 12K DNA Chip assay (Agilent, Santa Clara, USA) (p/n 5067-1508) was used to assess the fragment size distribution. Five µg of

sheared gDNA was DNA damage repaired and end-repaired using polishing enzymes. A blunt-end ligation reaction followed by exonuclease treatment was performed to create the SMRT bell template. A Blue Pippin device (Sage Science, Beverly, USA) was used to size-select the SMRT bell template and enrich large fragments >7 kbp. The sized-selected library was quality inspected and quantified on the Agilent Bioanalyzer 12 Kb DNA Chip and on a Qubit Fluorimeter, respectively. A ready to sequence SMRT bell-Polymerase Complex was created using the P6 DNA/Polymerase binding kit 2.0 (Pacific Biosciences) (p/n 100-236-500) according to the manufacturer's instructions. The Pacific Biosciences RS2 instrument was programmed to load and sequence the sample(s) on 9 SMRT cells v3.0 (Pacific Biosciences) (p/n100-171-800), taking 1 movie of 240 minutes each per SMRT cell. A MagBead loading (Pacific Biosciences) (p/n 100-133-600) method was chosen in order to improve the enrichment the longer fragments.

2.3.5 RNA sequencing

The quality of the isolated RNA was determined with a Qubit (1.0) Fluorometer and a Bioanalyzer 2100. Only those samples with a 260 nm/280 nm ratio between 1.8–2.1 and a 28S/18S ratio within 1.5–2 were further processed. The TruSeq RNA Sample Prep Kit v2 (Illumina Incorporation, San Diego, USA) was used in the succeeding steps. Briefly, total RNA samples (100-1000 ng) were poly A enriched and then reverse-transcribed into double-stranded cDNA. The cDNA samples were fragmented, end-repaired and polyadenylated before ligation of TruSeq adapters containing the index for multiplexing fragments containing TruSeq adapters on both ends were selectively enriched with PCR. The quality and quantity of the enriched libraries were validated using Qubit (1.0) Fluorometer and the Caliper GX LabChip GX (Caliper Life Sciences Incorporation, Waltham, USA). The product is a smear with an average fragment size of approximately 260 bp. The libraries were normalized to 10 nM in Tris-Cl 10 mM, pH 8.5 with 0.1% Tween 20.

The TruSeq PE Cluster Kit HS4000 or TruSeq SR Cluster Kit HS4000 (Illumina) was used for cluster generation using 10 pM of pooled normalized libraries on the cBOT. Sequencing was performed on the Illumina HiSeq 2000 paired end at 2 X101 bp or single end 100 bp using the TruSeq SBS Kit HS4000 (Illumina).

2.3.6 Data analyses

A detailed description of all the steps in the data analysis and of the methods used is available in the Supplementary Methods S1. Briefly, long fragments sequenced on the Pacific Bioscience RSII were trimmed and filtered. The final draft of the genome was obtained using Falcon and the annotation was performed using the Maker2 pipeline (English et al., 2012).

De novo transcriptome assembly was performed using the Trinity software suite v.2.2.0 (Grabherr et al., 2011) using default parameters followed by CDS-prediction with Transdecoder (http://transdecoder.sourceforge.net/) and cd-hit-est (http://weizhongli-lab.org/cd-hit/).

The RNA sequences were aligned to our annotated genome using STAR (Dobin et al., 2013), expression levels were quantified using the R package GenomicRanges (Lawrence et al., 2013) and differential expressions were tested for using the R package edgeR (Robinson et al., 2010).

Finally, Gene Ontology (GO) overrepresentation analyses were performed using PANTHER (Mi et al., 2017).

2.4 Results and Discussion

2.4.1 De novo genome assembly and annotation

The use of long reads to assemble the genome of *D. repens* resulted in 916 total contigs, representing an over 30-fold increase in compactness compared to the latest release of *D. immitis*

(www.nematodes.org). Similarly, a N50 of 584,065 bp equates to an over 50-fold increase in contiguity. The size of the assembled *D. repens* genome was just below 100 MB, around 17% larger than that of *D. immitis*, and their GC contents differed by only 0.7% (Table 1). In order for these comparisons to be appropriate, all the metrics in Table 1 refer to contigs. The annotation of the *de novo* genome assembly predicted 10,357 genes. This gene number is about 8.9% fewer than those predicted in *D. immitis,* and this difference is consistent at the transcript level with 11,262 transcripts predicted, about 8.8% fewer than those predicted in *D. immitis*. On average (median), *D. repens* shows a larger number of exons per gene (7 versus 5). At the same time, we report the exons being slightly shorter (136 vs. 142 bp, median size, Table 1). Such differences in the annotation could be partially reflecting the biological difference between the two nematodes and partially be a consequence of the fact that the *D. repens* transcriptome has been assembled from RNA samples extracted from microfilariae and L3 stages whereas that of *D. immitis* is based on male and female adult samples (Godel et al., 2012). Moreover, the fact that the annotation are less divergent than the assembly suggests that the most likely reason for such a difference in genome size lies in the repetitive regions that the all-short-reads assembly of *D. immitis* has not been able to resolve. By looking at some of the common quality metrics for a genome annotation, we find that more than 95% of the transcripts have an Annotation Edit Distance (AED, the main transcript quality score produced by MAKER2) below 0.5 (Supplementary Fig. 1) and over 60% of the transcripts were annotated to UniProtKB/Swiss-Prot. By comparing these values with the MAKER2 benchmarks (Campbell et al. 2014), we can state that the overall quality of our *D. repens* annotation is definitely high.

Of the 206 proteins that did not have an ortholog in the *D. immitis* transcriptome, 141 presented a homologous sequence in the genome of *D. immitis,* suggesting that they are indeed present in *D. immitis*, but that the current *D. immitis* annotation has not been able to produce a gene model

for them. Of the remaining 65, 16 could be annotated to UniProtKB/Swiss-Prot (Supplementary Table 1). Of particular interest could be the only protein out of the aforementioned 65 (augustus-000063F-processed-gene-0.33-mRNA-1) that, despite showing no ortholog either in the *D. immitis* transcriptome or genome, was found in all other considered filarial nematodes (*Brugyia malayi, B. pahangi, D. repens*, *L. loa, Onchocerca volvulus, Wuchereria bancrofti*) and *Caenorhabditis elegans*. A homolog search via blast to UniProtKB/Swiss-Prot revealed that the only match is in the organism *Plasmodium reichenowi* with the circumsporozoite protein, the immunodominant surface antigen on the sporozoite (the infective stage of the malaria parasite that is transmitted from the mosquito to the vertebrate host).

A putative set of 834 filarial-specific proteins has been reported (i.e. present in *D. immitis* and *B. malayi*, but absent in *C. elegans*, *Trichinella spiralis* and *Ascaris suum*) (Luck et al., 2014). Among the proteins identified in our annotation, 712 overlap with the aforementioned set, 448 of which are annotated to UniProtKB/Swiss-Prot. Of the remaining 122, 50 could be annotated to UniProtKB/Swiss-Prot (Supplementary Table 2). It is possible that the difference in the qualities of the assemblies plays a role, however, in both cases the N50 is well above the average gene length, and therefore the discrepancies at the annotation level, as expected, are much less severe. We would rather argue that this is simply a consequence of the fact that our *D. repens* annotation has overall fewer transcripts predicted than that of *D. immitis*.

Since the biology of *L. loa* is more similar to that of *D. repens* than it is to *D. immitis*, we checked whether some differences could be identified at the molecular level. More precisely, we compared the proportion of orthodox transcripts between these species. Since we appreciate that the sets of predicted proteins of the current annotations of these organisms might be incomplete, we first aligned the 6-frame translated genomes to the *L. loa* proteins, and this results in only around 15% ORFs finding a match. We then also used Blat to align the transcripts

against the genome. The latter resulted in over 80% of the transcripts finding a match. Interestingly, in both cases, we found that the *L. loa* proteome shows a significant enrichment ($p < 0.01$ and $p < 0.00005$, respectively) for *D.repens* proteins compared to *D.immitis* proteins. This similarity resembles the macroscopic traits associated to these nematodes. In particular, *L. loa* is endemic to Central Africa and causes ocular and systemic symptoms in humans. Similar to *D. repens*, adult *L. loa* are located in loose connective tissue beneath the skin and between the fascial layers on top of somatic muscles (Whittaker et al., 2018).

Supplementary Figure 2 shows the GO BP which are shared by *D. repens* and *L. loa*, but not present in *D. immitis*. Interestingly, many of the significant GO BP are motility-related (locomotion, regulation of locomotion) and feeding-related (regulation of pharyngeal pumping).

2.4.2 Gene expression analysis

The analysis of the RNA-sequencing data of mf and L3 revealed a clear separation of the two groups upon unsupervised clustering (Fig. 1A) and PCA analysis (Fig. 2). By scrutinising these results, one also notices that the intra-group analysis of the RNA-sequencing data revealed a lower variance among the mf samples than among the L3 samples. However, since the first principal component (inter-groups) accounts for a percentage of variance more than four times that of the second component (intra-groups), this effect can be considered small compared to the intra-group variance. Moreover, by looking at the intra-groups scatter plots, also showing a higher overall homogeneity of the group Mf (smaller dispersion), no particular sample is an outlier (Supplementary Fig. 3). Substantial differences between the groups were confirmed by the differential expression analysis, with 876 genes reported to be differentially expressed at thresholds of 0.05 for the FDR. Of these, 591 could be annotated to UniProtKB/Swiss-Prot. Cluster signatures for the upregulated genes in the two groups can be readily recognized when plotting the expression profiles in a heatmap (Fig. 1B).

For further discussion, only genes with an UniProt annotation to the filarial nematodes *D. immitis, B. malayi, B. pahangi, O. volvulus,* the model organism *C. elegans* or a UniProtKB/Swiss-Prot annotation to the endosymbiont bacteria *Wolbachia sp.* and *W. pipientis* have been considered.

In the L3 stage, 358 genes were upregulated compared to the mf stage. Out of these upregulated genes, 155 genes had a UniProtKB/Swiss-Prot annotation to *D. immitis, C. elegans, B. malayi, B. pahangi, O. volvulus,* and none to *Wolbachia* spp. Many of the highly transcribed genes were described as genes that are part of structural components (e.g. cuticle collagens (P34687, P18835, P18833), are involved in muscle development (O01761), muscle contraction and locomotion (P02566, Q11176, P19625, Q11176) or are genes that code for proteins used for hydrolysis and proteases (Q04457,O44451).

The identification of genes that are uniquely upregulated in the L3 stage provide potential targets for the development of improved diagnostic tools to screen mosquitoes for infectious stages by detecting these overexpressed genes based on cDNA. One gene (P21249) that was upregulated in the *D. repens* L3 stage is described as "major antigen" or "myosin-like antigen" (OVT1) in *O. volvulus*, a parasitic filarial nematode (https://www.uniprot.org/uniprot/P21249). There are two potential homologues of the OVT1 gene (nDi.2.2.2.t09053-RA and nDi.2.2.2.t01724-RA) in the heartworm *D. immitis*, which both are upregulated in the *D. immitis* L3 and the L4 stage compared to the mf (Luck et al., 2014). One of these genes (nDi.2.2.2.t01724-RA) is also upregulated in the L4 stage compared to the L3 stage, which indicates that these genes are likely to be stage markers. Thus, these genes seem to be promising targets for diagnostic PCR of infectious L3 stages in mosquitoes. However, final conclusions on the suitability of this target cannot be made yet, as deeper stage-specific analyses of the further mosquito-associated stages (sausage stage L1, L2) of *D. repens* are required. These stages are immobile, in contrast to the

highly motile L3, and cannot easily be recognized under the binocular after grinding the mosquitoes and cannot be separated from the Malpighian tubules. Though high quantities of RNA were extracted from samples containing sausage stage L1 or L2 at days 5 and 8 post inoculation, respectively, the quality was very low, probably consisting mainly of degraded mosquito RNA, and could therefore not be used for further analysis (data not shown).

In the mf stage, a total of 235 genes were upregulated compared to the L3-stage. Out of these, 57 genes had a UniProtKB/Swiss-Prot annotation to the same nematodes mentioned above and 8 to *Wolbachia* spp. Many of these upregulated genes are located in the cell nucleus and concern DNA (P04255) and RNA binding (Q09524), transcription (Q9NAL4), chemotaxis (Q8I7F8), and some genes are described as stress response genes (P29778,). In total, seven *Wolbachia* proteins were upregulated in the mf stage, including elongation factors (Q5GSU1, Q73H58) and chaperon proteins (B3CNB5, Q73I71) indicating high protein synthesis activity. Corresponding to the findings with *D. immitis*, upregulation of *Wolbachia* genes was only detected for mf stages (Luck et al., 2014). By mapping our transcripts to those of the study on *D. immitis* (Luck et al., 2014), we identified 59 and 3 genes reported to be significantly upregulated in both studies the L3 and mf groups, respectively (Supplementary Table 3) and the effect sizes for these genes are remarkably comparable (Supplementary Fig. 4). Many more such genes were identified (254) in that study, however a non-negligible part of those probably represent false positives associated with the absence of biological replicates (n=1).

In order to perform exploratory gene-set analyses, we used a subset of 153 genes which were upregulated in the L3 group and mapped to the model organism *C. elegans* and the subset of 57 genes which were upregulated in the mf group and mapped to *C. elegans* to interrogate the GO databases and the Reactome Pathway database using Panther. After a gene enrichment analysis, several significantly enriched terms were found in all the databases for the genes upregulated in the L3 group (Fig. 3), while only a handful of GO Biological Processes and one

generic pathway were reported for the mf-stage unregulated genes (Fig. 4). The significantly enriched GO BP for the L3 group could be classified into 6 terms (Fig. 3). These findings match with the biology of the L3 of *D. repens,* which have to migrate from the Malpighian tubules to the mosquito vector's mouthparts (proboscis) and penetrate the host's skin at the bite site (Simon et al., 2012). Soon after the infection, the L3 moult to the L4. For this, the L3 have to be motile and contract their muscles and change the anatomical structure (prepare for moulting and morphogenesis of the anatomical structures). This is supported by the fact that also the GO CC and GO MF as well as the reactome which are significantly enriched are involved in muscle contractions and related metabolisms.

For the mf group, only 3 significantly enriched GO BP could be identified with the gene enrichment analysis (Fig. 4), which are basically associated to cellular component organization or biogenesis and ammonium ion metabolic process. Microfilariae represent the first developmental stage of *D. repens* preparing for life in a different host (invertebrate vs. vertebrate), a phase which requires many metabolic changes such as the development of internal organs. This may explain the upregulation of many GO terms involving metabolic processes in this life stage.

In order to add further depth to the potentially limited statistical power of gene-set analysis when short input genes lists are used, we also simply looked at the distribution of the genes across the Panther protein classes (Fig. 5). For the third larval stage, the largest number of genes that was upregulated belonged to the class of hydrolases and cytoskeletal proteins (Fig. 5A), and in the mf stage to genes encoding proteins for nucleic acid binding, but also for cytoskeletal proteins which can be explained as both larval stages undergo morphological changes and are under development.

In the present study, parasite material from three different European countries was used (adult worm from Italy for de novo genome sequencing, microfilariae from Lithuania, L3 produced

with material from Croatia). However, these different origins of the samples should not have biased our results, as an earlier study had revealed only very small genetic differenced in *D. repens* isolates from Europe (Yilmaz et al., 2016).

3.4.3 Wolbachia

The assembly of the endosymbiont *Wolbachia* (see Supplementary Methods S1 for the details) resulted in 44 contigs, predicting a total size for the bacterial genome of 818,820 bp which is in the same size range as the genome size of *Wolbachia* of *D. immitis* (0.92 MB) and *B. malayi* (1.08 Mb) (Godel et al., 2012). *Loa loa* does not have *Wolbachia* endosymbionts (McGarry et al., 2003). *Wolbachia* are classified into supergroups based on three genes, ftsZ, groEL and gltA (Casiraghi et al., 2005). *Wolbachia* from *D. immitis* and *O. volvulus* for example belong to the supergroup C, while those from *B. malayi* and *W. bancrofti* belong to the supergroup D.

The phylogenetic trees based on the homologues of these genes in our *D. repens* assembly are in full agreement with those in a previous study (Casiraghi et al., 2005), i.e. *Wolbachia* from *D. repens* clusters with those from *D. immitis* (Fig. 6).

3.4.4 Conclusions

In this study, we present the first genome assembly draft and annotation of *D. repens*, a filarial nematode causing subcutaneous and ocular infections in dogs and cats. By sequencing high-coverage Pacbio long reads, we have been able to generate very compact and contiguous assemblies for both *D. repens* and its endosymbiont *Wolbachia*, for which a full genome has also been missing so far. When compared to the only other *Dirofilaria* for which such data are available (*D. immitis*), the quality metric of our assembly are orders of magnitudes higher, making our dataset a valuable resource for more specific, in depth studies of Dirofilariae.

Moreover, the comparative analysis of the mf and third-larva groups provides a snapshot of the molecular differences between these two developmental stages and a first step to better understand the mechanisms behind the specific abilities of these filarial nematodes to interact in different ways with different hosts and to ensure the completion of their life cycles.

Genome and transcriptome data of nematodes were exploited previously to identify novel intervention methods (drug and vaccine targets) and host-parasite interactions of filarial nematodes (McNulty et al., 2016; Gasser et al., 2017; Grote et al., 2017; Bennuru et al., 2018). At the same time, it is important to stay critical with the interpretation of newly gathered information about protein-coding genes. The advancements in genome sequencing have also led to an accumulation of hypothetical proteins, and results could get compared with possibly miss-annotated genes since many helminth annotations are still based on "primary-sequence-level search protocols" and, additionally, different methods are used to deal with hypothetical proteins (Palevich et al., 2018).

2.5 References

Ballesteros, C., Tritten, L., O'Neill, M., Burkman, E., Zaky, W.I., Xia, J., Moorhead, A., Williams, S.A., Geary, T.G., 2016. The effect of in vitro cultivation on the transcriptome of adult *Brugia malayi*. PLOS Neglect. Trop. Dis. 10, e0004311.

Bennuru, S., O'Connell, E.M., Drame, P.M., Nutman, T.B., 2018. Mining filarial genomes for diagnostics and therapeutic targets. Trends Parasitol. 34, 80-90.

Campbell, M. S., Holt, C., Moore, B. and Yandell, M. 2014. Genome annotation and curation using MAKER and MAKER-P. Curr. Protoc. Bioinformatics 48:4.11.1-4.11.39

Capelli, G., Genchi, C., Baneth, G., Bourdeau, P., Brianti, E., Cardoso, L., Danesi, P., Fuehrer, H-P., Gianelli, A., Ionica, A.M., Maia, C., Modry, D., Montarsi, F., Krücken, J., Papadopoulos, E., Petric, D., Pfeffer, M., Savic, S., Otranto, D., Poppert, S., Silaghi, C., 2018. Recent advances on *Dirofilaria repens* in dogs and humans in Europe. Parasite Vector, 11, 663.

Casiraghi, M., Bordenstein, S.R., Baldo, L., Lo, N., Beninati, T., Wernegreen, J.J., Werren, J.H., Bandi, C., 2005. Phylogeny of *Wolbachia pipientis* based on gltA, groEL and ftsZ gene sequences: clustering of arthropod and nematode symbionts in the F supergroup, and evidence for further diversity in the *Wolbachia* tree. Microbiology 151, 4015-4022.

Chaisson, M.J., Tesler, G., 2012. Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory. BMC Bioinformatics 13, 238.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., Gingeras, T.R., 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15-21.

English, A.C., Richards, S., Han, Y., Wang, M., Vee, V., Qu, J., Qin, X., Muzny, D.M., Reid, J.G., Worley, K.C., Gibbs, R.A., 2012. Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing technology. PLOS ONE 7, e47768.

Gasser, R.B., Tan, P., Teh, B.T., Wongkham, S., Young, N.D., 2017. Genomics of worms, with an emphasis on *Opisthorchis viverrini* - opportunities for fundamental discovery and biomedical outcomes. Parasitol. Int. 66, 341-345.

Genchi, C., Genchi, M., Petry, G., Kruedewagen, E.M., Schaper, R., 2013. Evaluation of the efficacy of imidacloprid 10% / moxidectin 2.5% (Advocate®, Advantage® Multi, Bayer) for the prevention of *Dirofilaria repens* infection in dogs. Parasitol. Res. 112 Suppl 1, 81-89.

Genchi, C., Kramer, L., 2017. Subcutaneous dirofilariosis (*Dirofilaria repens*): an infection spreading throughout the old world. Parasite Vector 10, 517.

Genchi, C., Mortarino, M., Rinaldi, L., Cringoli, G., Traldi, G., Genchi, M., 2011. Changing climate and changing vector-borne disease distribution: the example of *Dirofilaria* in Europe. Vet. Parasitol. 176, 295-299.

Godel, C., Kumar, S., Koutsovoulos, G., Ludin, P., Nilsson, D., Comandatore, F., Wrobel, N., Thompson, M., Schmid, C.D., Goto, S., Bringaud, F., Wolstenholme, A., Bandi, C., Epe, C., Kaminsky, R., Blaxter, M., Maser, P., 2012. The genome of the heartworm, *Dirofilaria immitis*, reveals drug and vaccine targets. FASEB J. 26, 4650-4661.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29, 644-652.

Grote, A., Lustigman, S., Ghedin, E., 2017. Lessons from the genomes and transcriptomes of filarial nematodes. Mol. Biochem. Parasitol. 215, 23-29.

Lawrence, M., Huber, W., Pages, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M.T., Carey, V.J., 2013. Software for computing and annotating genomic ranges. PLOS Computational Biology 9, e1003118.

Luck, A.N., Evans, C.C., Riggs, M.D., Foster, J.M., Moorhead, A.R., Slatko, B.E., Michalski, M.L., 2014. Concurrent transcriptional profiling of *Dirofilaria immitis* and its *Wolbachia* endosymbiont throughout the nematode life cycle reveals coordinated gene expression. BMC Genomics 15, 1041.

Magi, M., Calderini, P., Gabrielli, S., Dell'Omodarme, M., Macchioni, F., Prati, M.C., Cancrini, G., 2008. *Vulpes vulpes*: a possible wild reservoir for zoonotic filariae. Vector Borne Zoonotic Dis. 8, 249-252.

McGarry, H.F., Pfarr, K., Egerton, G., Hoerauf, A., Akue, J.-P., Enyong, P., Wanji, S., Klager, S.L., Bianco, A.E., Beeching, N.J., Taylor, M.J., 2003. Evidence against *Wolbachia* symbiosis in *Loa loa*. Filaria J. 2, 9.

McNulty, S.N., Strube, C., Rosa, B.A., Martin, J.C., Tyagi, R., Choi, Y.J., Wang, Q., Hallsworth Pepin, K., Zhang, X., Ozersky, P., Wilson, R.K., Sternberg, P.W., Gasser, R.B., Mitreva, M., 2016. *Dictyocaulus viviparus* genome, variome and transcriptome elucidate lungworm biology and support future intervention. Sci. Rep. 6, 20316.

Mi, H., Huang, X., Muruganujan, A., Tang, H., Mills, C., Kang, D., Thomas, P.D., 2017. PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. Nucleic Acids Res. 45, D183-D189.

Palevich, N., Britton, C., Kamenetzky, L., Mitreva, M., de Moraes Mourao, M., Bennuru, S., Quack, T., Scholte, L.L.S., Tyagi, R., Slatko, B.E., 2018. Tackling hypotheticals in helminth genomes. Trends Parasitol. 34, 179-183.

Rishniw, M., Barr, S.C., Simpson, K.W., Frongillo, M.F., Franz, M., Dominguez Alpizar, J.L., 2006. Discrimination between six species of canine microfilariae by a single polymerase chain reaction. Vet. Parasitol. 135, 303-314.

Robinson, M.D., McCarthy, D.J., Smyth, G.K., 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139-140.

Silaghi, C., Beck, R., Capelli, G., Montarsi, F., Mathis, A., 2017. Development of *Dirofilaria immitis* and *Dirofilaria repens* in *Aedes japonicus* and *Aedes geniculatus*. Parasite Vector 10, 94.

Simon, F., Siles-Lucas, M., Morchon, R., Gonzalez-Miguel, J., Mellado, I., Carreton, E., Montoya-Alonso, J.A., 2012. Human and animal dirofilariasis: the emergence of a zoonotic mosaic. Clin. Microbiol. Rev. 25, 507-544.

Slatko, B.E., Taylor, M.J., Foster, J.M., 2010. The *Wolbachia* endosymbiont as an anti-filarial nematode target. Symbiosis 51, 55-65.

Taylor, M.J., Bandi, C., Hoerauf, A., 2005. *Wolbachia* bacterial endosymbionts of filarial nematodes. Adv. Parasitol. 60, 245-284.

Whittaker, C., Walker, M., Pion, S.D.S., Chesnais, C.B., Boussinesq, M., Basanez, M.-G., 2018. The Population biology and transmission dynamics of *Loa loa*. Trends Parasitol. 34, 335-350.

Yilmaz, E., Fritzenwanker, M., Pantchev, N., Lendner, M., Wongkamchai, S., Otranto, D., Kroidl, I., Dennebaum, M., Le, T.H., Anh Le, T., Ramunke, S., Schaper, R., von Samson-

Himmelstjerna, G., Poppert, S., Krucken, J., 2016. The mitochondrial genomes of the zoonotic canine filarial parasites *Dirofilaria* (*Nochtiella*) *repens* and *Candidatus* Dirofilaria (Nochtiella) Honkongensis provide evidence for presence of cryptic species. PLOS Negl. Trop. Dis. 10, e0005028.

2.6 Tables

Table 1. Summary of some relevant metrics associated to the assembly and annotation of the *Dirofilaria repens* and other filarial nematodes

Metric	D. repens(*)	D. immitis(**)	B. Malayi(**)	L. Loa(**)	C. Elegans(***)
Assembly size (Mb)	99.59	78.16	87.95	96.36	100.27
Number of contigs	916	11'654	205	2'183	6
contig N50 (bp)	584'065	15'962	10'384'967	180'288	17,493,829
Largest contig (bp)	4'800'701	195'591	23'878'448	1'570'872	20'924'180
Protein-coding gene models	10'357	11'375	10'959	12'473	20'362
Predicted proteins	11'262	12'344	15'393	12'473	32'061
Protein-coding sequence (%)	15.5	18.0	13.8	5.4	25.4
Annotated genes	6'467	8'113	NA	8'592	10'190
Annotated transcripts	7'044	8'113	NA	8'592	17'862
Overall GC content (%)	27.6	28.3	28.5	30.8	35.4
Median exons per gene	7	5	5	9	6
Median exon size	136	142	139	138	147
(*) Source: present study					
(**) Source: https://parasite.wormbase.org					
(***) Source: https://wormbase.org					

2.7 Figures

Figure 1. Hierarchical clustering of the *Dirofilaria repens* genes and samples based on gene expression data using the R function *hclust*. A) Unsupervised clustering based on the top 500 genes as ranked by variance across all the samples. B) Clustering based on the genes with FDR < 0.05 in the differential expression analysis.

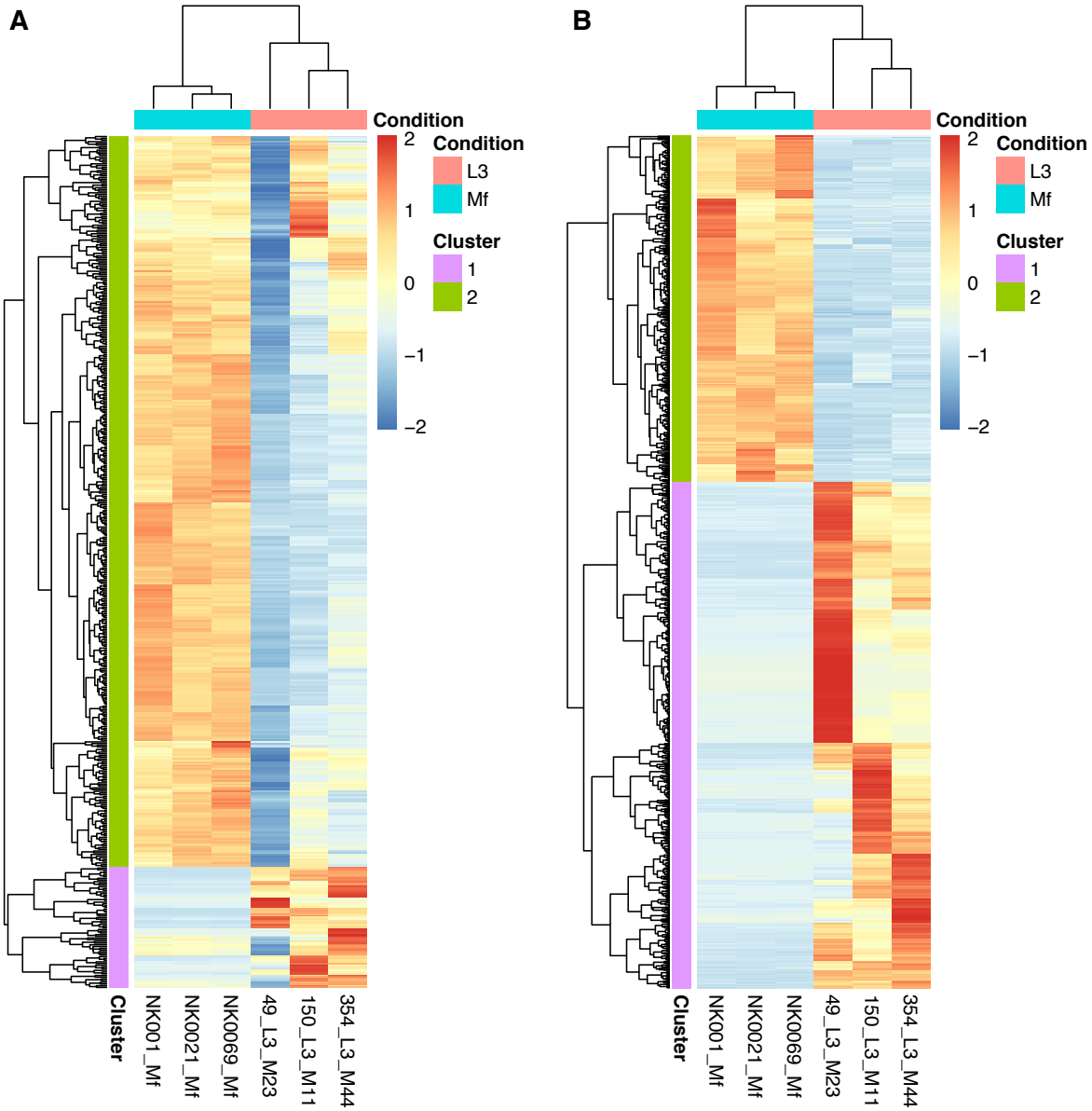


Figure 2. Principal component analysis based on the top 500 *Dirofilaria repens* genes as ranked by variance across all the samples. The number in the axis levels is the percentage of variation explained by the two main components.

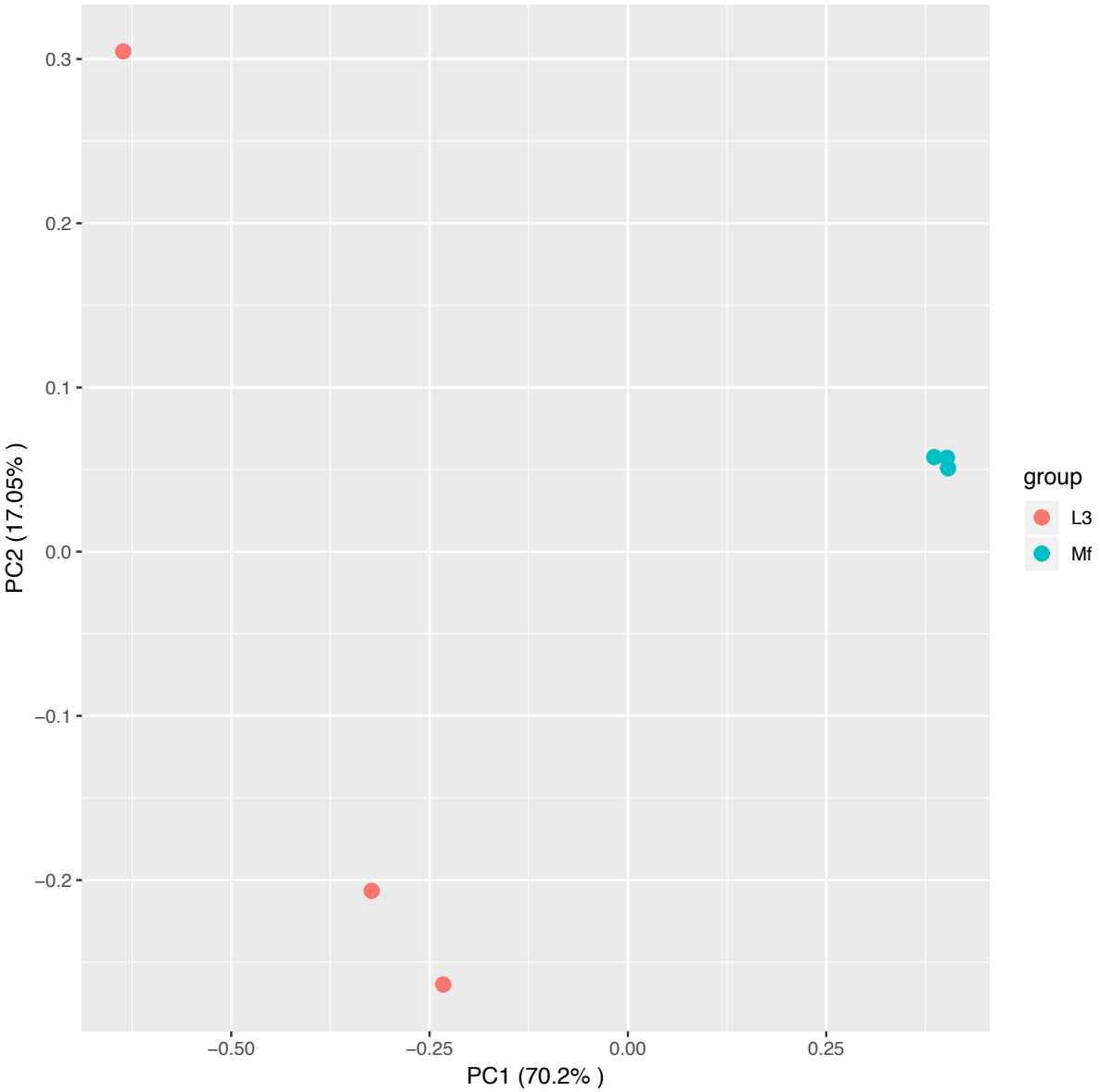


Figure 3. List of GO databases and Reactome pathway terms with an enrichment analysis FDR < 0.05 based on the genes upregulated in the *Dirofilaria repens* L3 stage.

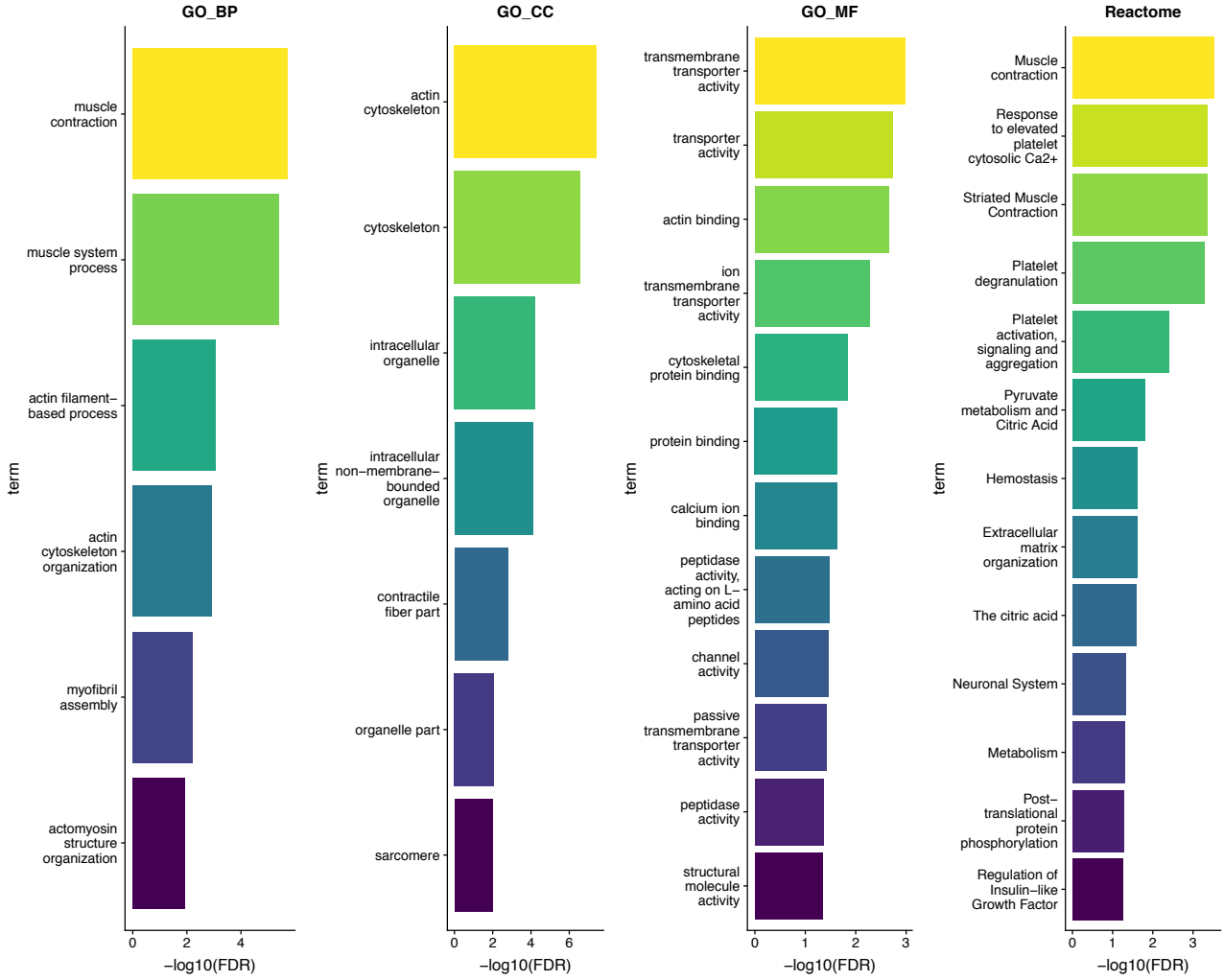


Figure 4. List of GO databases and Reactome pathway terms with an enrichment analysis FDR < 0.05 based on the genes upregulated in the *Dirofilaria repens* mf stage.

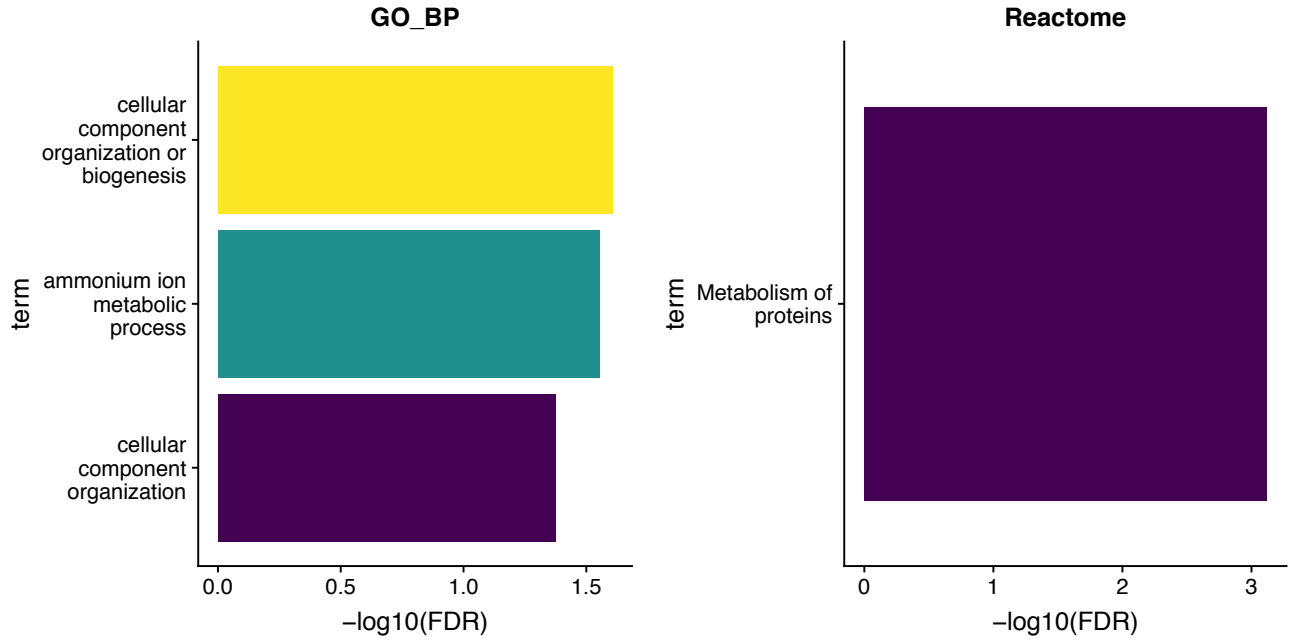


Figure 5. Distribution of the differentially regulated genes in the various Panther protein classes. A) genes upregulated in the *Dirofilaria repens* L3 stage. B) genes upregulated in the mf stage.

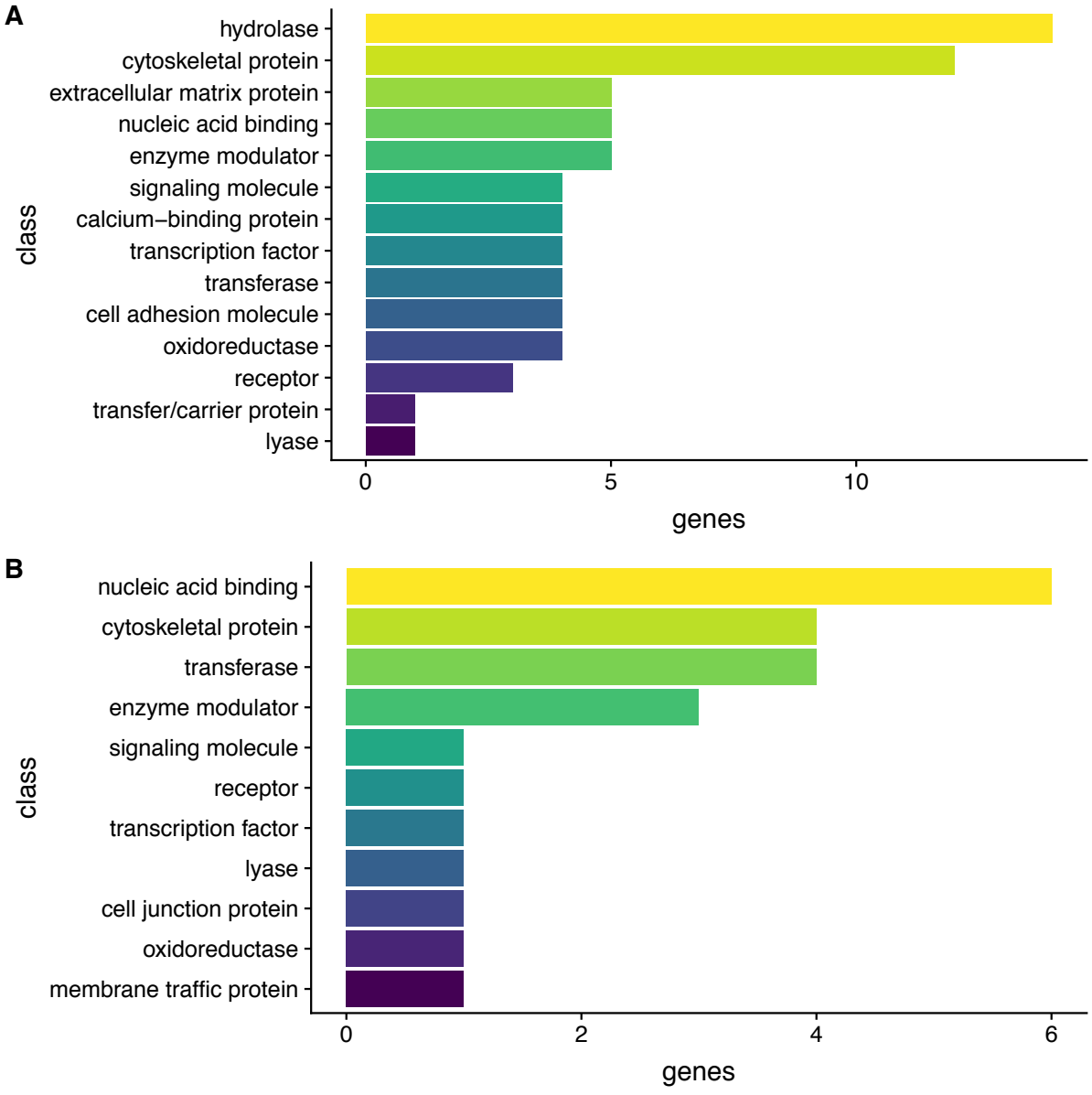
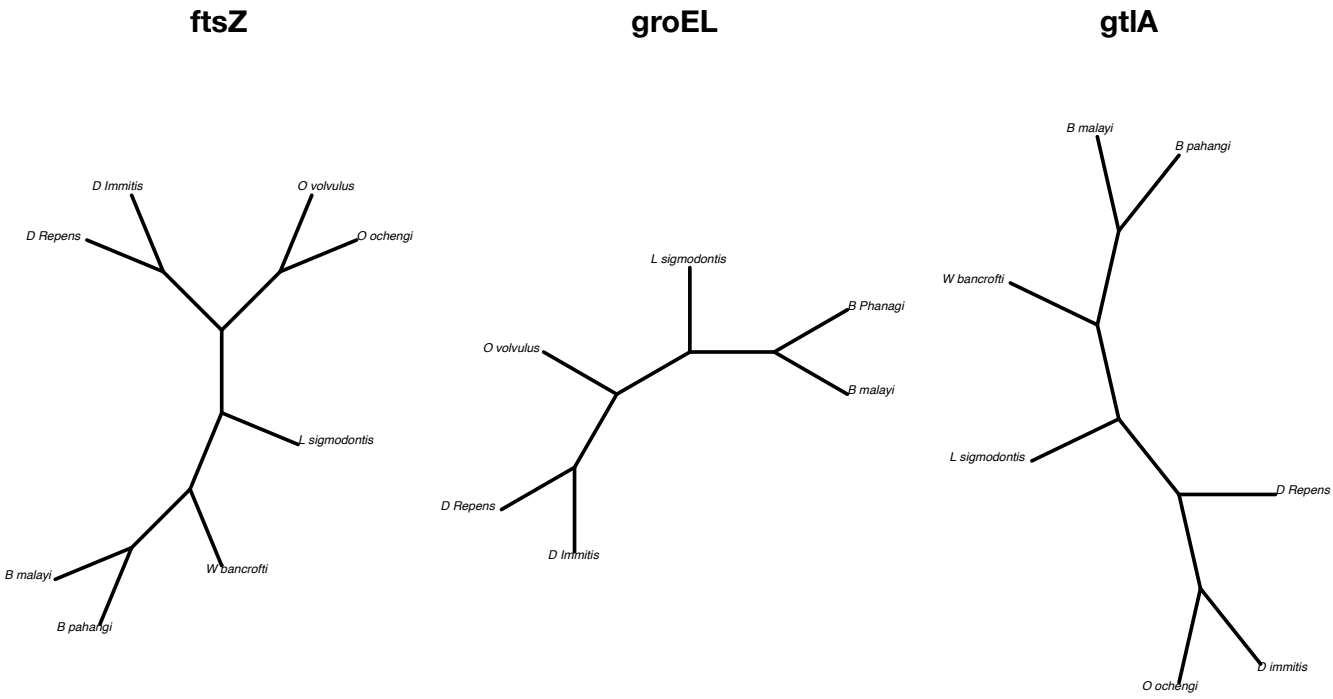


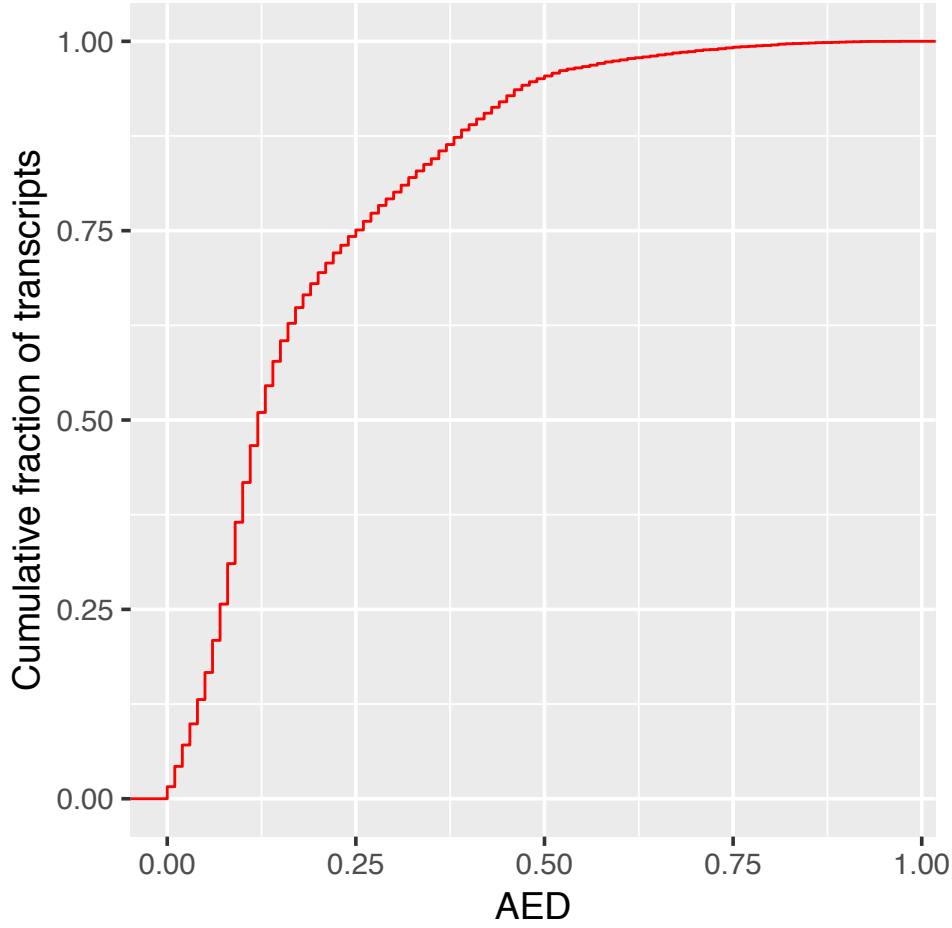
Figure 6. Phylogenetic trees of the filarial *Wolbachia* based on the sequences of the *ftsZ*, *groEL* and *gtlA* genes.



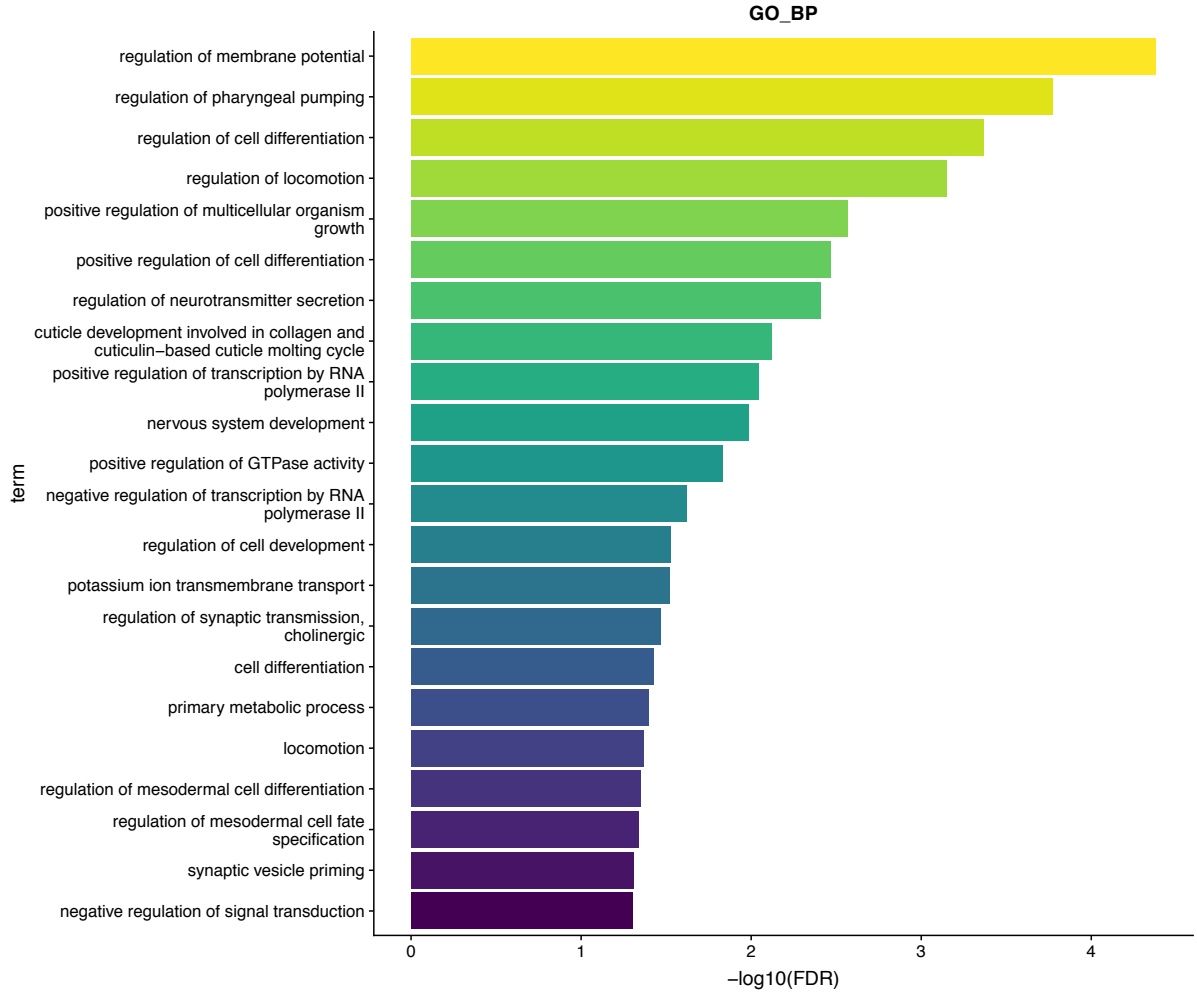
3. Anhänge

3.1. Supplementary Figures

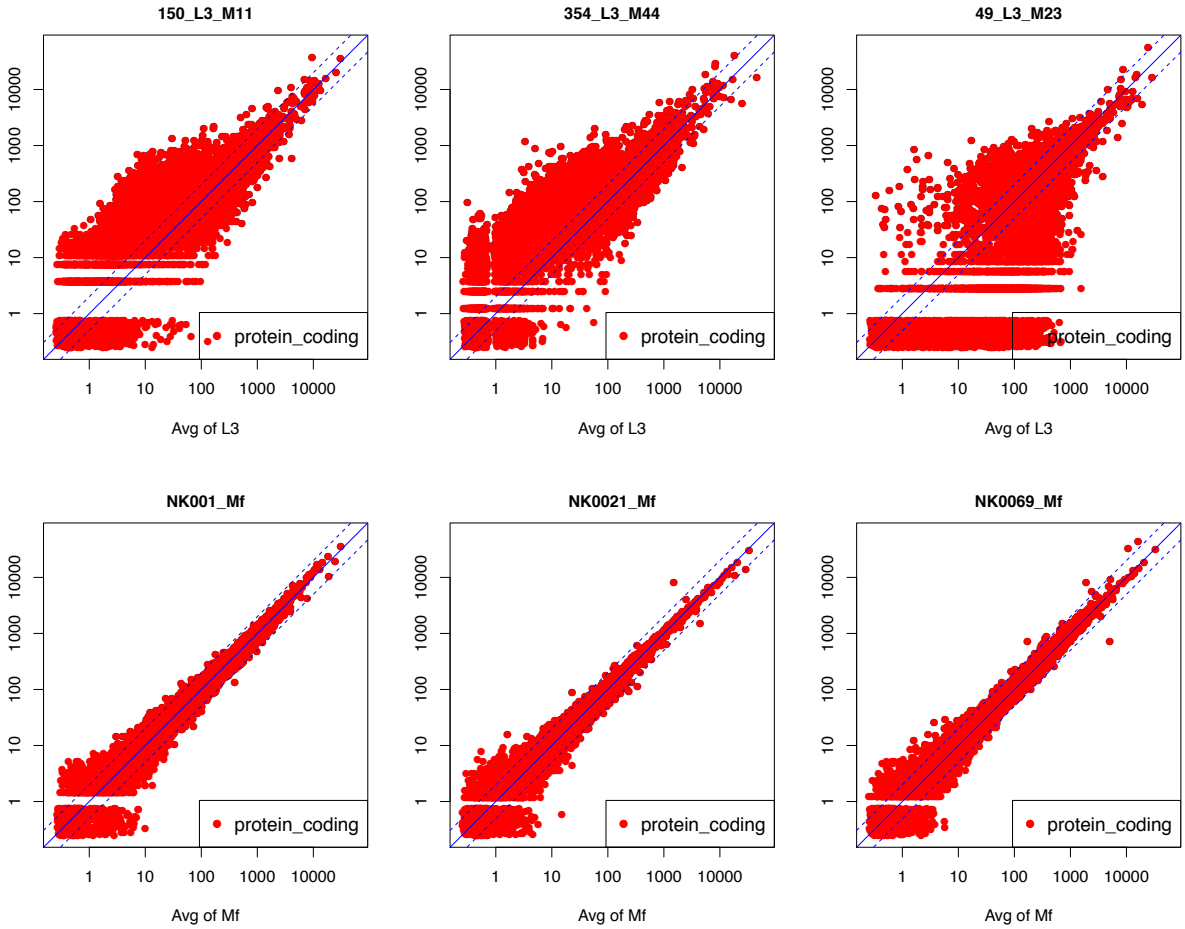
Supplementary Figure 1. Cumulative Annotation Edit Distance (AED) scores for the annotated transcripts in our *Dirofilaria repens* annotation.



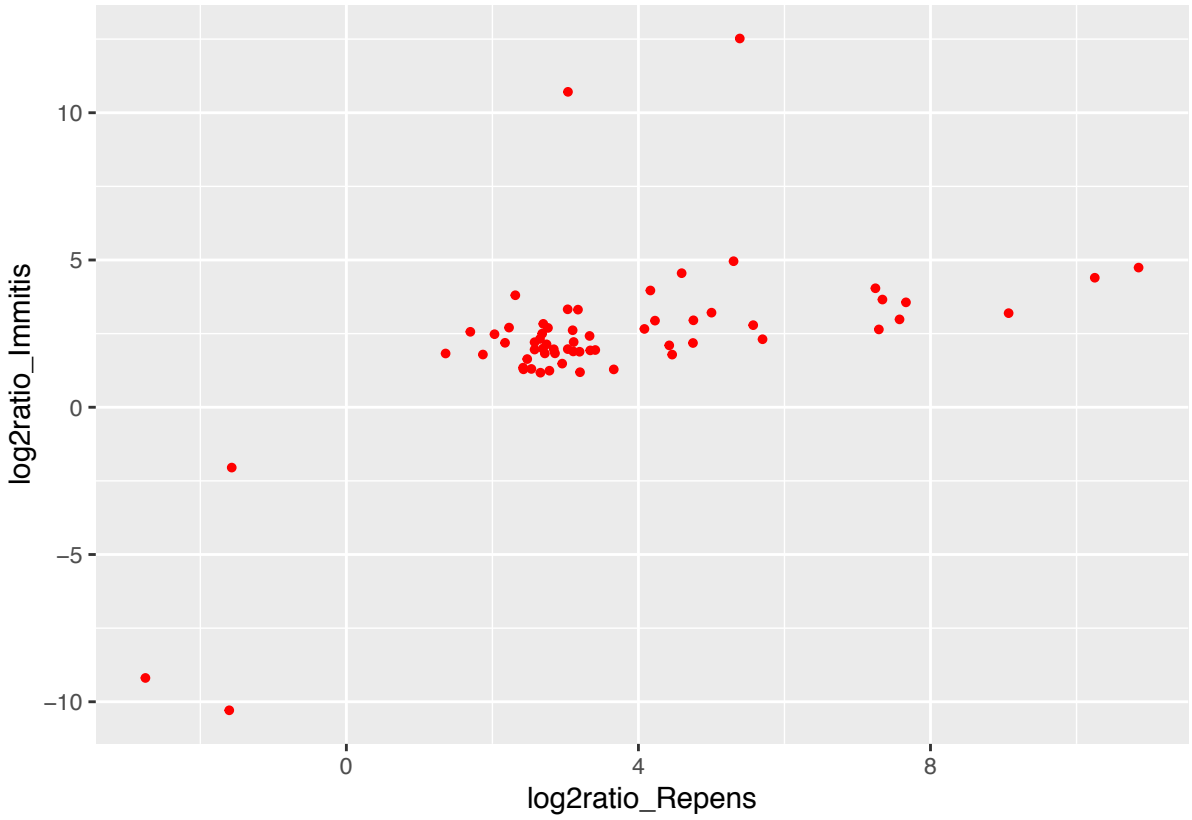
Supplementary Figure 2. List of GO Biological Processes with an enrichment analysis FDR <0.05 based on the transcripts shared by *Dirofilaria repens* and *Loa loa* but not present in *D. immitis*



Supplementary Figure 3. Intra-groups individual dispersion plots for all the samples. In each sample, for each gene (red dot), the TMM normalized expression value is plotted against the average in the group.



Supplementary Figure 4. Scatterplot of the log2-ratios in our study and in the study of Luck et al. (2014) for the genes which in both studies have been reported to be differentially regulated between stages L3 and mf.



3.2. Supplementary Tables

Supplementary Table 1: List of 16 annotated proteins identified in *Dirofilaria repens* but for which no ortholog could be found either in the proteome or in the genome of *D. immitis*. This protein-protein search has been performed using diamond blasts, while the transcripts-genome search using Blat.

protID	annotation	organism
augustus-000059F-processed-gene-0.42-mRNA-1	sp|Q73HP2|GATA_WOLPM_Glutamyl-tRNA(Gln)_amidotransferase_subunit_A_OS=Wolbachia_pipientis_wMel_GN=gatA_PE=3_SV=1	Wolbachia_pipientis_wMel
augustus-000059F-processed-gene-0.51-mRNA-1	sp|P73633|Y1875_SYNY3_Uncharacterized_protein_slr1875_OS=Synechocystis_sp._(strain_PCC_6803_/_Kazusa)_GN=slr1875_PE=4_SV=1	Synechocystis_sp._(strain_PCC_6803_/_Kazusa)
augustus-000059F-processed-gene-0.69-mRNA-1	sp|P45485|FTSZ_WOLSP_Cell_division_protein_FtsZ_OS=Wolbachia_sp._GN=ftsZ_PE=3_SV=1	Wolbachia_sp.
augustus-000059F-processed-gene-1.79-mRNA-1	sp|Q8CRJ7|ZDH1_STAES_Zinc-type_alcohol_dehydrogenase-like_protein_SE_1777_OS=Staphylococcus_epidermidis_(strain_ATCC_12228)_GN=SE_1777_PE=3_SV=1	Staphylococcus_epidermidis_(strain_ATCC_12228)
augustus-000107F-processed-gene-0.103-mRNA-1	sp|P57061|LOLC_NEIMA_Lipoprotein-releasing_system_transmembrane_protein_LolC_OS=Neisseria_meningitidis_serogroup_A_/_serotype_4A_(strain_Z2491)_GN=lolC_PE=3_SV=1	Neisseria_meningitidis_serogroup_A_/_serotype_4A_(strain_Z2491)
augustus-000107F-processed-gene-0.19-mRNA-1	sp|Q73I13|DAPB_WOLPM_4-hydroxy-tetrahydrodipicolinate_reductase_OS=Wolbachia_pipientis_wMel_GN=dapB_PE=3_SV=1	Wolbachia_pipientis_wMel
augustus-000107F-processed-gene-0.40-mRNA-1	sp|Q73FY9|NDK_WOLPM_Nucleoside_diphosphate_kinase_OS=Wolbachia_pipientis_wMel_GN=ndk_PE=3_SV=1	Wolbachia_pipientis_wMel
augustus-000107F-processed-gene-0.80-mRNA-1	sp|P61479|HSLV_WOLPM_ATP-dependent_protease_subunit_HslV_OS=Wolbachia_pipientis_wMel_GN=hslV_PE=3_SV=1	Wolbachia_pipientis_wMel
augustus-000174F-processed-gene-0.15-mRNA-1	sp|Q73IZ0|DNAA_WOLPM_Chromosomal_replication_initiator_protein_DnaA_OS=Wolbachia_pipientis_wMel_GN=dnaA_PE=3_SV=1	Wolbachia_pipientis_wMel
augustus-000190F-processed-gene-0.20-mRNA-1	sp|B3CLI3|MNMG_WOLPP_tRNA_uridine_5-carboxymethylaminomethyl_modification_enzyme_MnmG_OS=Wolbachia_pipientis_subsp._Culex_pipiens_(strain_wPip)_GN=mnmG_PE=3_SV=1	Wolbachia_pipientis_subsp._Culex_pipiens_(strain_wPip)
augustus-000190F-processed-gene-0.35-mRNA-1	sp|P43746|DPO3X_HAEIN_DNA_polymerase_III_subunit_tau/gamma_OS=Haemophilus_influenzae_(strain_ATCC_51907_/_DSM_11121_/_KW20_/_Rd)_GN=dnaX_PE=3_SV=1	Haemophilus_influenzae_(strain_ATCC_51907_/_DSM_11121_/_KW20_/_Rd)
maker-000014F-augustus-gene-4.20-mRNA-1	sp|Q61LC0|S35B3_CAEBR_Adenosine_3'-phospho_5'-phosphosulfate_transporter_2_OS=Caenorhabditis_briggsae_GN=pst-2_PE=3_SV=1	Caenorhabditis_briggsae
maker-000059F-augustus-gene-0.0-mRNA-1	sp|Q73IJ6|DEF_WOLPM_Peptide_deformylase_OS=Wolbachia_pipientis_wMel_GN=def_PE=3_SV=1	Wolbachia_pipientis_wMel
maker-000059F-augustus-gene-1.4-mRNA-1	sp|Q5GTQ0|MNME_WOLTR_tRNA_modification_GTPase_MnmE_OS=Wolbachia_sp._subsp._Brugia_malayi_(strain_TRS)_GN=mnmE_PE=3_SV=1	Wolbachia_sp._subsp._Brugia_malayi_(strain_TRS)
maker-000059F-augustus-gene-1.4-mRNA-1	sp|Q5GTQ0|MNME_WOLTR_tRNA_modification_GTPase_MnmE_OS=Wolbachia_sp._subsp._Brugia_malayi_(strain_TRS)_GN=mnmE_PE=3_SV=1	Wolbachia_sp._subsp._Brugia_malayi_(strain_TRS)
maker-000107F-augustus-gene-0.1-mRNA-1	sp|Q5GSD6|RS6_WOLTR_30S_ribosomal_protein_S6_OS=Wolbachia_sp._subsp._Brugia_malayi_(strain_TRS)_GN=rpsF_PE=3_SV=1	Wolbachia_sp._subsp._Brugia_malayi_(strain_TRS)

Supplementary Table 2. List of 50 annotated filarial-specific proteins (i.e., identified in *Dirofilaria immitis* and *Brugia malayi*, but absent in *Caenorhabditis elegans*, *Trichinella spiralis* and *Ascaris suum)* or which no ortholog could be found in *D. immitis*.

***D.Immitis* protein ID**	**annotation**
nDi.2.2.2.g00373	von willebrand factor type a domain containing protein
nDi.2.2.2.g00374	nose resistant to fluoxetine protein 6
nDi.2.2.2.g00375	nuclear hormone receptor family member nhr-41
nDi.2.2.2.g00376	histidine acid phosphatase family protein
nDi.2.2.2.g00377	taf3 protein
nDi.2.2.2.g00378	lectin c-type domain-containing protein
nDi.2.2.2.g00379	nicotinic acetylcholine receptor alpha subunit 8
nDi.2.2.2.g00380	pdz domain containing protein
nDi.2.2.2.g00381	cuticle collagen dpy-7
nDi.2.2.2.g00382	wd-repeat protein 23
nDi.2.2.2.g00383	tk fer protein kinase
nDi.2.2.2.g00384	calcium ding protein homolog family member (cbn-1)
nDi.2.2.2.g00385	phd finger protein 10
nDi.2.2.2.g00386	cutical collagen 6
nDi.2.2.2.g00387	collagen alpha-1 chain
nDi.2.2.2.g00388	ste20-like serine threonine-protein kinase
nDi.2.2.2.g00389	leucine rich repeat family protein
nDi.2.2.2.g00390	moz sas family protein
nDi.2.2.2.g00391	calreticulin family protein
nDi.2.2.2.g00392	hypothetical protein LOAG_00413 [Loa loa]
nDi.2.2.2.g00393	serpin
nDi.2.2.2.g00394	ground-like domain containing protein
nDi.2.2.2.g00395	sodium potassium-transporting atpase subunit alpha-1
nDi.2.2.2.g00396	mitogen-activated protein kinase kinase kinase
nDi.2.2.2.g00397	basement membrane proteoglycan
nDi.2.2.2.g00398	kelch-like protein 10-like
nDi.2.2.2.g00399	beta-glucuronidase precursor
nDi.2.2.2.g00400	cell death specification protein 2
nDi.2.2.2.g00401	guanine nucleotide-binding protein g subunit alpha
nDi.2.2.2.g00402	uncoordinated family member (unc-2)
nDi.2.2.2.g00403	bromodomain protein
nDi.2.2.2.g00404	protein kinase domain containing protein
nDi.2.2.2.g00405	vesicle-fusing atpase
nDi.2.2.2.g00406	cre-unc-15 protein
nDi.2.2.2.g00407	glutamate decarboxylase
nDi.2.2.2.g00408	dys-1 protein
nDi.2.2.2.g00409	hypothetical protein Bm1_53085 [Brugia malayi]
nDi.2.2.2.g00410	aldolase
nDi.2.2.2.g00411	uncoordinated family member (unc-62)
nDi.2.2.2.g00412	progastricsin (pepsinogen c)
nDi.2.2.2.g00413	animal heme peroxidase
nDi.2.2.2.g00414	adenylate kinase 1
nDi.2.2.2.g00415	fructosamine-3-kinase-related protein
nDi.2.2.2.g00416	bardet-biedl syndrome 2 protein homolog
nDi.2.2.2.g00417	cmgc mapk jnk protein kinase
nDi.2.2.2.g12029	vacuolar amino acid transporter 4
nDi.2.2.2.g12147	cre-cgef-1 protein
nDi.2.2.2.g12250	leucine-rich repeat protein soc-2 homolog
nDi.2.2.2.g12316	chitin synthase 2 (chs-2) fragment

Supplementary Table 3. Annotated list of the 62 proteins also identified as differentially regulated in a previous study[1]

1 1. Luck, A.N., Evans, C.C., Riggs, M.D., Foster, J.M., Moorhead, A.R., Slatko, B.E., Michalski, M.L., 2014. Concurrent transcriptional profiling of *Dirofilaria immitis* and its *Wolbachia* endosymbiont throughout the nematode life cycle reveals coordinated gene expression. BMC Genomics 15, 1041.

D_Repens_ID	D_Immitis_ID	annotation	log2ratio_Repens	log2ratio_Immitis	upregulation
augustus-000000F-processed-gene-0_12	nDi.2.2.2.g05855	sp|P48812|G3P_BRUMA_Glyceraldehyde-3-phosphate_dehydrogenase_OS=Brugia_malayi_GN=G3PD_PE=2_SV=1	3.113	2.216	L3
augustus-000000F-processed-gene-23_1	nDi.2.2.2.g03196	sp|P36573|LEC1_CAEEL_32_kDa_beta-galactoside-binding_lectin_OS=Caenorhabditis_elegans_GN=lec-1_PE=1_SV=1	2.857	1.831	L3
augustus-000001F-processed-gene-0_46	nDi.2.2.2.g05819	sp|Q10651|A4_CAEEL_Beta-amyloid-like_protein_OS=Caenorhabditis_elegans_GN=apl-1_PE=1_SV=2	2.477	1.637	L3
augustus-000001F-processed-gene-3_20	nDi.2.2.2.g02834	sp|Q23551|UNC22_CAEEL_Twitchin_OS=Caenorhabditis_elegans_GN=unc-22_PE=1_SV=3	1.359	1.825	L3
augustus-000002F-processed-gene-4_51	nDi.2.2.2.g01594	sp|Q21253|GELS1_CAEEL_Gelsolin-like_protein_1_OS=Caenorhabditis_elegans_GN=gsnl-1_PE=1_SV=1	3.032	1.975	L3
augustus-000005F-processed-gene-1_36	nDi.2.2.2.g01923	sp|P20631|COL13_CAEEL_Cuticle_collagen_13_OS=Caenorhabditis_elegans_GN=col-13_PE=2_SV=1	4.592	4.552	L3
augustus-000007F-processed-gene-8_15	nDi.2.2.2.g10005	sp|Q20140|KAD1_CAEEL_Probable_adenylate_kinase_isoenzyme_F38B2.4_OS=Caenorhabditis_elegans_GN=F38B2.4_PE=2_SV=1	2.652	2.320	L3
augustus-000013F-processed-gene-1_29	nDi.2.2.2.g01186	sp|Q09201|SFXN1_CAEEL_Putative_sideroflexin-1.1_OS=Caenorhabditis_elegans_GN=sfxn-1.1_PE=2_SV=1	4.227	2.941	L3
augustus-000028F-processed-gene-2_3	nDi.2.2.2.g09053	sp|P21249|ANT1_ONCVO_Major_antigen_OS=Onchocerca_volvulus_GN=OVT1_PE=2_SV=2	7.293	2.642	L3
augustus-000063F-processed-gene-0_19	nDi.2.2.2.g10543	sp|Q9XUY5|AT1B3_CAEEL_Probable_sodium/potassium-transporting_ATPase_subunit_beta-3_OS=Caenorhabditis_elegans_GN=nkb-3_PE=2_SV=1	2.426	1.288	L3
augustus-000066F-processed-gene-0_79	nDi.2.2.2.g08235	sp|P22085|CYTX_ONCVO_Onchocystatin_OS=Onchocerca_volvulus_PE=2_SV=2	4.164	3.965	L3
augustus-000074F-processed-gene-0_71	nDi.2.2.2.g08675	sp|P90850|YCF2E_CAEEL_Uncharacterized_peptidase_C1-like_protein_F26E4.3_OS=Caenorhabditis_elegans_GN=F26E4.3_PE=1_SV=3	4.082	2.659	L3
augustus-000092F-processed-gene-0_82	nDi.2.2.2.g07247	sp|P15796|H12_CAEEL_Histone_H1.2_OS=Caenorhabditis_elegans_GN=hil-2_PE=1_SV=3	3.031	3.326	L3
augustus-000097F-processed-gene-0_28	nDi.2.2.2.g02466	sp|Q1ENI8|PXDN_CAEEL_Peroxidasin_homolog_OS=Caenorhabditis_elegans_GN=pxn-1_PE=1_SV=1	2.956	1.481	L3
augustus-000110F-processed-gene-0_16	nDi.2.2.2.g09994	sp|P02566|MYO4_CAEEL_Myosin-4_OS=Caenorhabditis_elegans_GN=unc-54_PE=4_SV=1	2.173	2.188	L3
augustus-000133F-processed-gene-0_14	nDi.2.2.2.g10488	sp|P34804|COL40_CAEEL_Cuticle_collagen_40_OS=Caenorhabditis_elegans_GN=col-40_PE=2_SV=3	5.303	4.959	L3
augustus-000133F-processed-gene-0_24	nDi.2.2.2.g09009	sp|Q10901|EAA1_CAEEL_Excitatory_amino_acid_transporter_OS=Caenorhabditis_elegans_GN=glt-1_PE=1_SV=2	1.698	2.562	L3
maker-000000F-augustus-gene-20_10	nDi.2.2.2.g00484	sp|Q09665|TNNC2_CAEEL_Troponin_C,_isoform_2_OS=Caenorhabditis_elegans_GN=tnc-2_PE=2_SV=1	3.035	10.709	L3
maker-000000F-augustus-gene-6_39	nDi.2.2.2.g02937	sp|Q27371|TNNT_CAEEL_Troponin_T_OS=Caenorhabditis_elegans_GN=mup-2_PE=2_SV=1	5.001	3.212	L3
maker-000001F-augustus-gene-12_34	nDi.2.2.2.g01724	sp|P21249|ANT1_ONCVO_Major_antigen_OS=Onchocerca_volvulus_GN=OVT1_PE=2_SV=2	2.781	1.241	L3
maker-000001F-augustus-gene-18_42	nDi.2.2.2.g03741	sp|Q11176|WDR1_CAEEL_Actin-interacting_protein_1_OS=Caenorhabditis_elegans_GN=unc-78_PE=1_SV=1	2.534	1.301	L3
maker-000001F-augustus-gene-8_8	nDi.2.2.2.g02333	sp|P19625|MLR1_CAEEL_Myosin_regulatory_light_chain_1_OS=Caenorhabditis_elegans_GN=mlc-1_PE=2_SV=1	2.228	2.705	L3
maker-000002F-augustus-gene-12_25	nDi.2.2.2.g01862	sp|P21249|ANT1_ONCVO_Major_antigen_OS=Onchocerca_volvulus_GN=OVT1_PE=2_SV=2	4.462	1.788	L3
maker-000003F-augustus-gene-1_41	nDi.2.2.2.g02051	sp|Q09476|PXL1_CAEEL_Paxillin_homolog_1_OS=Caenorhabditis_elegans_GN=pxl-1_PE=2_SV=2	3.409	1.944	L3
maker-000003F-augustus-gene-11_21	nDi.2.2.2.g00413	sp|P17657|DPY13_CAEEL_Cuticle_collagen_dpy-13_OS=Caenorhabditis_elegans_GN=dpy-13_PE=1_SV=2	5.572	2.787	L3
maker-000003F-augustus-gene-11_22	nDi.2.2.2.g00414	sp|P17656|COL2_CAEEL_Cuticle_collagen_2_OS=Caenorhabditis_elegans_GN=col-2_PE=2_SV=1	4.422	2.101	L3
maker-000003F-augustus-gene-3_23	nDi.2.2.2.g04976	sp|P90901|IFA1_CAEEL_Intermediate_filament_protein_ifa-1_OS=Caenorhabditis_elegans_GN=ifa-1_PE=1_SV=2	4.753	2.952	L3
maker-000003F-augustus-gene-6_10	nDi.2.2.2.g00663	sp|P34328|HSP10_CAEEL_Heat_shock_protein_Hsp-12.2_OS=Caenorhabditis_elegans_GN=hsp-12.2_PE=3_SV=1	4.747	2.181	L3
maker-000005F-augustus-gene-5_8	nDi.2.2.2.g00886	sp|Q17745|TRXR1_CAEEL_Thioredoxin_reductase_1_OS=Caenorhabditis_elegans_GN=trxr-1_PE=2_SV=3	2.658	1.173	L3
maker-000006F-augustus-gene-8_1	nDi.2.2.2.g02405	sp|Q09665|TNNC2_CAEEL_Troponin_C,_isoform_2_OS=Caenorhabditis_elegans_GN=tnc-2_PE=2_SV=1	2.314	3.803	L3
maker-000007F-augustus-gene-1_18	nDi.2.2.2.g08295	sp|Q09508|DHSA_CAEEL_Succinate_dehydrogenase_[ubiquinone]_flavoprotein_subunit,_mitochondrial_OS=Caenorhabditis_elegans_GN=sdha-1_PE=1_SV=3	2.841	1.969	L3
maker-000007F-augustus-gene-7_0	nDi.2.2.2.g08467	sp|P34687|COL34_CAEEL_Cuticle_collagen_34_OS=Caenorhabditis_elegans_GN=col-34_PE=2_SV=2	7.664	3.562	L3
maker-000008F-augustus-gene-0_24	nDi.2.2.2.g06547	sp|O01761|UNC89_CAEEL_Muscle_M-line_assembly_protein_unc-89_OS=Caenorhabditis_elegans_GN=unc-89_PE=1_SV=3	1.870	1.789	L3
maker-000013F-augustus-gene-0_13	nDi.2.2.2.g01192	sp|Q9XUN9|TNNI3_CAEEL_Troponin_I_3_OS=Caenorhabditis_elegans_GN=tni-3_PE=2_SV=1	3.340	1.931	L3
maker-000013F-augustus-gene-4_17	nDi.2.2.2.g03607	sp|Q18066|DIM_CAEEL_Disorganized_muscle_protein_1_OS=Caenorhabditis_elegans_GN=dim-1_PE=1_SV=3	2.682	2.499	L3
maker-000015F-augustus-gene-4_49	nDi.2.2.2.g04267	sp|Q21752|VDAC_CAEEL_Probable_voltage-dependent_anion-selective_channel_OS=Caenorhabditis_elegans_GN=R05G6.7_PE=2_SV=2	2.421	1.344	L3
maker-000015F-augustus-gene-4_50	nDi.2.2.2.g04268	sp|P34714|SPRC_CAEEL_SPARC_OS=Caenorhabditis_elegans_GN=ost-1_PE=1_SV=1	3.662	1.285	L3
maker-000019F-augustus-gene-1_0	nDi.2.2.2.g01527	sp|Q09254|HAR1_CAEEL_Hemiasterlin_resistant_protein_1_OS=Caenorhabditis_elegans_GN=har-1_PE=2_SV=1	2.718	1.828	L3
maker-000020F-augustus-gene-2_46	nDi.2.2.2.g02805	sp|Q09614|PTC1_CAEEL_Protein_patched_homolog_1_OS=Caenorhabditis_elegans_GN=ptc-1_PE=1_SV=2	3.200	1.190	L3
maker-000030F-augustus-gene-1_1	nDi.2.2.2.g09661	sp|P29030|CHIT_BRUMA_Endochitinase_OS=Brugia_malayi_PE=1_SV=1	7.576	2.983	L3
maker-000048F-augustus-gene-1_42	nDi.2.2.2.g09186	sp|P34575|CISY_CAEEL_Probable_citrate_synthase,_mitochondrial_OS=Caenorhabditis_elegans_GN=cts-1_PE=1_SV=1	3.331	2.422	L3
maker-000049F-augustus-gene-1_12	nDi.2.2.2.g11450	sp|Q09321|YR51_CAEEL_Uncharacterized_protein_F42A8.1_OS=Caenorhabditis_elegans_GN=F42A8.1_PE=2_SV=1	5.388	12.521	L3
maker-000056F-augustus-gene-1_24	nDi.2.2.2.g07823	sp|P54216|ALF1_CAEEL_Fructose-bisphosphate_aldolase_1_OS=Caenorhabditis_elegans_GN=aldo-1_PE=1_SV=1	3.194	1.884	L3
maker-000062F-augustus-gene-1_0	nDi.2.2.2.g10472	sp|P34687|COL34_CAEEL_Cuticle_collagen_34_OS=Caenorhabditis_elegans_GN=col-34_PE=2_SV=2	5.700	2.309	L3
maker-000063F-augustus-gene-1_20	nDi.2.2.2.g08791	sp|O17473|CATL_BRUPA_Cathepsin_L-like_OS=Brugia_pahangi_PE=1_SV=1	9.070	3.195	L3
maker-000079F-augustus-gene-0_72	nDi.2.2.2.g10091	sp|P52717|YUW5_CAEEL_Uncharacterized_serine_carboxypeptidase_F41C3.5_OS=Caenorhabditis_elegans_GN=F41C3.5_PE=1_SV=1	7.342	3.657	L3
maker-000085F-augustus-gene-0_20	nDi.2.2.2.g05468	sp|Q06561|UNC52_CAEEL_Basement_membrane_proteoglycan_OS=Caenorhabditis_elegans_GN=unc-52_PE=1_SV=2	3.099	2.615	L3
maker-000112F-augustus-gene-0_21	nDi.2.2.2.g03332	sp|P53014|MLE_CAEEL_Myosin,_essential_light_chain_OS=Caenorhabditis_elegans_GN=mlc-3_PE=1_SV=1	2.697	2.831	L3
maker-000118F-augustus-gene-0_47	nDi.2.2.2.g11375	sp|P55955|TTR16_CAEEL_Transthyretin-like_protein_16_OS=Caenorhabditis_elegans_GN=ttr-16_PE=1_SV=2	2.691	1.989	L3
maker-000126F-augustus-gene-0_30	nDi.2.2.2.g05499	sp|Q02171|MYSP_ONCVO_Paramyosin_OS=Onchocerca_volvulus_PE=2_SV=1	2.578	1.959	L3
maker-000126F-augustus-gene-0_30	nDi.2.2.2.g10698	sp|Q02171|MYSP_ONCVO_Paramyosin_OS=Onchocerca_volvulus_PE=2_SV=1	2.578	2.209	L3
maker-000129F-augustus-gene-0_27	nDi.2.2.2.g11900	sp|O17433|1CPX_DIRIM_1-Cys_peroxiredoxin_OS=Dirofilaria_immitis_PE=2_SV=1	2.029	2.480	L3
maker-000129F-augustus-gene-0_35	nDi.2.2.2.g06308	sp|P46426|GSTP_DIRIM_Glutathione_S-transferase_OS=Dirofilaria_immitis_PE=2_SV=1	2.742	2.136	L3
maker-000135F-augustus-gene-0_24	nDi.2.2.2.g09190	sp|Q10657|TPIS_CAEEL_Triosephosphate_isomerase_OS=Caenorhabditis_elegans_GN=tpi-1_PE=1_SV=2	3.172	3.314	L3
maker-000142F-augustus-gene-0_28	nDi.2.2.2.g00856	sp|P34687|COL34_CAEEL_Cuticle_collagen_34_OS=Caenorhabditis_elegans_GN=col-34_PE=2_SV=2	10.850	4.742	L3
maker-000144F-augustus-gene-0_20	nDi.2.2.2.g08684	sp|P18835|COL19_CAEEL_Cuticle_collagen_19_OS=Caenorhabditis_elegans_GN=col-19_PE=2_SV=2	10.250	4.398	L3
maker-000149F-augustus-gene-0_29	nDi.2.2.2.g06639	sp|P18834|COL14_CAEEL_Cuticle_collagen_14_OS=Caenorhabditis_elegans_GN=col-14_PE=2_SV=2	7.246	4.041	L3
maker-000184F-augustus-gene-0_10	nDi.2.2.2.g07893	sp|Q03575|TTR5_CAEEL_Transthyretin-like_protein_5_OS=Caenorhabditis_elegans_GN=ttr-5_PE=2_SV=1	3.107	1.895	L3
maker-000189F-augustus-gene-0_7	nDi.2.2.2.g07227	sp|Q07749|ADF2_CAEEL_Actin-depolymerizing_factor_2,_isoform_c_OS=Caenorhabditis_elegans_GN=unc-60_PE=1_SV=1	2.761	2.694	L3
maker-000038F-augustus-gene-1_8	nDi.2.2.2.g10723	sp|O44712|AHR_CAEEL_Aryl_hydrocarbon_receptor_protein_1_OS=Caenorhabditis_elegans_GN=ahr-1_PE=1_SV=1	-1.604	-10.290	mf
maker-000058F-augustus-gene-0_14	nDi.2.2.2.g03952	sp|P34688|DPY7_CAEEL_Cuticle_collagen_dpy-7_OS=Caenorhabditis_elegans_GN=dpy-7_PE=1_SV=1	-2.752	-9.191	mf
maker-000071F-augustus-gene-0_83	nDi.2.2.2.g05692	sp|Q21955|MED15_CAEEL_Mediator_of_RNA_polymerase_II_transcription_subunit_15_OS=Caenorhabditis_elegans_GN=mdt-15_PE=1_SV=3	-1.570	-2.049	mf

Supplementary Table 4. NCBI BioProjects ID of the protein databases used in this study for homology search.

Organism	NCBI BioProject Number
Caenorhabditis elegans	PRJNA13758
Litomosoides Sigmodontis	PRJEB3075
Wuchereria Bancrofti	PRJEB536
Loa Loa	PRJNA246086
Brugia Pahanagi	PRJEB497
Brugia Malayi	PRJNA10729
Onchocerca Ochengi	PRJEB1809
Onchocerca Volvulus	PRJEB513

3. 3 Supplementary Methods

3.3.1 Genome sequencing, assembly and annotation

Long fragments sequenced on the Pacific Bioscience RSII had the SMRTbell adapters removed and filtered according to a minimum read length of 50 nucleotides and a minimum read quality of 75%. The sample has been sequenced on 9 SMRT Cells, for a total of 792,532 reads (88,059 +/- 4264 per cell), averaging a length 11,654 bases. At 9,269.4 Megabases (1030 +/- 147.9 Mi) the throughput represents just about 100x coverage of the expected 100Mi bp total genome size (*D. repens* and *Wolbachia*). The proportions were maintained after splitting the reads, with about 130Mi bp mapping to *Wolbachia* and 9,122.2 remaining for the assembly of *D. repens*.

The final draft of the genome was obtained using Falcon (https://github.com/PacificBiosciences/FALCON/) with settings optimized for a medium sized genome (Supplementary table 1). Genome annotation was performed using the Maker2 pipeline (English et al., 2012) with the following options: a) protein sequences from *D. immitis* (www.nematodes.org, release nDi v2.2.2) were used as protein evidence; b) the transcripts generated by the de novo assembly pipeline were used as EST evidence; c) the *D. immitis* GenBank transcript file (www.nematodes.org, release nDi v2.2.2) was used to train the etraining Augustus algorithm (Keller et al., 2011) and the resulting models for the species were used as ab initio gene predictions. All intra-species sequence similarity searches at the transcripts or amino acid level were performed using DIAMOND (Buchfink et al., 2015). Homologous sequences of the transcripts in the genome were searched using Blat (Kent, 2002) with a e-value threshold of 0.01. The protein database used for the annotation was UniProtKB/Swiss-Prot. All the other protein resources are described in Supplementary Table 4.

3.3.2 De novo transcriptome assembly

De novo transcriptome assembly was performed using the Trinity software suite v.2.2.0 (Grabherr et al., 2011) using default parameters. The likely coding DNA sequences (CDS) and corresponding proteins within the de novo transcriptome assembly were predicted with Transdecoder (http://transdecoder.sourceforge.net/) following the "Running TransDecoder" procedure described at https://github.com/TransDecoder/TransDecoder/wiki and using the output of a homology search with D.Immitis as homology options. Potential redundancy was further decreased by applying the clustering-based method cd-hit-est with default settings (http://weizhongli-lab.org/cd-hit/). The set of transcripts finally obtained was used to infer gene prediction in the annotation of the *D. repens* genome.

3.3.3 Differential expression analysis

Despite the overall high quality of our assembly, the outcome still needs to be considered a draft. Consequently, in order to increase the robustness of the analysis of the RNA-sequencing data, the transcript expressions were quantified with two different methods. First, sequencing adapters were removed with Trimmomatic (Bolger et al., 2014). Secondly, the RNA sequences were aligned to our annotated genome using STAR (Dobin et al., 2013) and the distribution of the reads across genomic isoforms and their expression levels were quantified using the R package GenomicRanges (Lawrence et al., 2013) from Bioconductor Version 3.0. Differential expressions were tested using the R package edgeR (Robinson et al., 2010) from Bioconductor Version 3.0 by using the raw counts as input and the TMM normalisation method. Significantly differentially regulated genes that fulfilled the following two conditions, were retained: 1) Genes with a raw count value of 10 in at least two samples of at least one group. 2) genes with an fdr (Benjamini-Hochberg) lower than 0.05.

3.3.4 Gene-set analysis

Genes-set analysis were performed using Panther (http://pantherdb.org). As functional analyses are heavily organism-dependent, only the genes mapping to Caenorhabditis elegans were considered. Gene Ontology (GO) overrepresentation analyses were performed using PANTHER (Mi et al., 2017). The PANTHER GO-slim databases, Panther protein class and Reactome were the databases used for the GO, protein class and pathway analyses, respectively. In the overrepresentation test, categories which showed a Bonferroni-corrected enrichment p-value below 0.05 were reported as significant.

3.3.5 Wolbachia

The final draft of the genome was obtained using Falcon (https://github.com/PacificBiosciences/FALCON/) with settings optimized for a small sized genome. The phylogenetic trees were based on the gene sequences from the different netmatodes corresponding to the accession numbers in a previous study (Casiraghi et al., 2005). In order to extract the candidate sequences from our Wolbachia assembly, putative ORFfinder (https://www.ncbi.nlm.nih.gov/orffinder/) were predicted and blasted against the ftsZ and groEL from *D.repens* and gtlA from *D.Immitis*. The resulting sequences were multiply aligned to those from the other available nematodes using the R package *msa* (https://bioconductor.org/packages/release/bioc/html/msa.html). Finally, the Phylogenetic trees based on specific gene sequences were built using R packages *msa* (Bodenhofer et sl., 2015) and *ape (Paradis and Schliep, 2019)*.

3.3.6 References

Bodenhofer U, Bonatesta E, Horejs-Kainrath C, Hochreiter S 2015. msa: an R package for multiple sequence alignment. Bioinformatics, 31(24), 3997–3999.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114-2120.

Buchfink, B., Xie, C., Huson, D.H., 2015. Fast and sensitive protein alignment using DIAMOND. Nat. Meth. 12, 59-60.

Casiraghi, M., Bordenstein, S.R., Baldo, L., Lo, N., Beninati, T., Wernegreen, J.J., Werren, J.H., Bandi, C., 2005. Phylogeny of *Wolbachia pipientis* based on gltA, groEL and ftsZ gene sequences: clustering of arthropod and nematode symbionts in the F supergroup, and evidence for further diversity in the *Wolbachia* tree. Microbiology 151, 4015-4022.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., Gingeras, T.R., 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15-21.

English, A.C., Richards, S., Han, Y., Wang, M., Vee, V., Qu, J., Qin, X., Muzny, D.M., Reid, J.G., Worley, K.C., Gibbs, R.A., 2012. Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing technology. PLOS ONE 7, e47768.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29, 644-652.

Keller, O., Kollmar, M., Stanke, M., Waack, S. 2011. A novel hybrid gene prediction method employing protein multiple sequence alignments. Bioinformatics, 6, 757–763.

Kent, W.J. 2002. BLAT - the BLAST-like alignment tool. Genome Res. 12(4),656-64.

Lawrence, M., Huber, W., Pages, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M.T., Carey, V.J., 2013. Software for computing and annotating genomic ranges. PLOS Computational Biology 9, e1003118.

Mi, H., Huang, X., Muruganujan, A., Tang, H., Mills, C., Kang, D., Thomas, P.D., 2017. PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. Nucleic Acids Res. 45, D183-D189.

Paradis E. , Schliep, K., 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35, 526-528.

Robinson, M.D., McCarthy, D.J., Smyth, G.K., 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139-140.

4. Danksagung

Acknowledgments

We thank Relja Beck (Croatian Veterinary Institute, Zagreb, Croatia) and Mindaugas Šarkūnas (Lithuanian University of Health Sciences, Kaunas, Lithuania) for providing parasite material; Mathieu Blaser (formerly Novartis Animal Health, St. Aubin, Switzerland) for his advices and help regarding the mosquito laboratory colony; Chandra Ramakrishnan, Lucienne Tritten and Saša Štefanić (Institute of Parasitology, University of Zürich) for invaluable help in laboratory and data analyses work; Sirisha Aluri (Functional Genomics Center Zurich, Zürich, Switzerland) for advice on RNA extraction and transcriptome sequencing. The financial contributions of the Federal Food Safety and Veterinary Office (as sponsor of the National Centre for Vector Entomology) and Bayer Animal Health are highly acknowledged.

Competing interests

The authors declare that they have not competing interests.

Curriculum Vitae

Vorname Name	Chiara Valentina Cafarelli
Geburtsdatum	08.11.1989
Geburtsort	Zürich, Schweiz
Nationalität	Schweiz
Heimatort	Zweisimmen, Bern

08/1996– 07/2002	**Grundschule 1.-6. Klasse** (Schulhaus Fluntern, Zürich und Schulhaus Feldmeilen, Feldmeilen ZH, Schweiz)
09/2002– 06/2008	**Langzeitgymnasium** (Realgymnasium Rämibühl, Zürich, Schweiz, Neusprachliches Profil mit kleinem Latinum)
26.08.2008	**Matura** (Realgymnasium Rämibühl, Zürich, Schweiz)
09/2009 – 12/2016	**Studium der Veterinärmedizin** (Universität Zürich, Vetsuisse-Fakultät, Schweiz)
30.12.2016	**Abschlussprüfung vet. med.,** (Universität Zürich, Vetsuisse-Fakultät, Schweiz)
04/2016 – 04/2019	**Anfertigung der Dissertation** unter Leitung von Prof. Dr. med. vet. Cornelia Silaghi am Institut für Parasitologie der Vetsuisse-Fakultät Universität Zürich Direktor: Prof. Dr. med. vet. Peter Deplazes
09/2018 – heute	**Assistenzärztin Kleintiermedizin** bei Marigin AG – Zentrum für Tiermedizin, Feusisberg SZ, Schweiz